VARIATIONAL MONTE CARLO ESTIMATION OF THE DISSOCIATION ENERGY OF CuH USING CORRELATED SAMPLING.

By

Andrej Doboš

B. Sc. (Physics) Faculty of Mathematics and Physics of Comenius University in Bratislava, Slovakia

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER'S OF SCIENCE

in

THE FACULTY OF MATHEMATICS AND SCIENCE

DEPARTMENT OF PHYSICS

We accept this thesis as conforming

to the required standard

BROCK UNIVERSITY

September 1994

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the Brock University, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Physics
Brock University
500 Glenridge Avenue
St. Catharines, Ontario
Canada L2S 3A1

Date:

Abstract

A new approach to treating large Z systems by quantum Monte Carlo has been developed. It naturally leads to notion of the 'valence energy'. Possibilities of the new approach has been explored by optimizing the wave function for CuH and Cu and computing dissociation energy and dipole moment of CuH using variational Monte Carlo. The dissociation energy obtained is about 40% smaller than the experimental value; the method is comparable with SCF and simple pseudopotential calculations. The dipole moment differs from the best theoretical estimate by about 50% what is again comparable with other methods (Complete Active Space SCF and pseudopotential methods).

Table of Contents

Abstract ii

Acknowledgement viii

1 Introduction 1

2 Ab Initio Methods 3

2.1 Hartree-Fock Self-Consistent-Field Method 4

2.2 Configuration Interaction in SCF 6

2.3 Complete Active Space SCF 7

2.4 Monte Carlo Methods 8

3 Variational Monte Carlo 11

3.1 Monte Carlo Integration 11

3.2 The Generalized Metropolis Algorithm 13

3.3 Choice of the Transition Probability Matrix 14

3.4 Algorithm 17

3.5 Critical Slowdown and Split-τ Technique 19

3.6 Nodal Problems 21

3.6.1 Variable-τ Algorithm 21

4 $f \times g$ Approach to D_e 23

4.1 Inspiration 23

4.2 Basic Formulas 26

4.3 Separation of the Valence and Core Energy 28
4.4 Philosophy of the Program 29
4.5 The Variance Term 30
4.5.1 Origin of the Variance Term 30
4.5.2 Evaluation of the Variance Term 31
4.6 The Exact Solution 32
4.7 Choice of the Function f 33
4.8 Choice of the Function g 34
4.9 Estimate of the Variance Term 35
4.10 Optimization 37
4.11 Results 38
4.12 Results—Improved Sampling 45

5 Comments, Suggestions, and Conclusions **47**
5.1 Configuration Interaction Approach 47
5.2 Valence Energy 48
5.3 Core Penetration 49
5.4 Damped-Core Quantum MC 51
5.5 Conclusions 52

Appendices **54**

A Diffusion Quantum Monte Carlo **54**
A.1 Basic Idea 54
A.2 Fermion Case 57

B Energy and Dipole Moment in $f \times g$ Approach **59**
B.1 Potential Energy 59

B.2 Kinetic Energy 60
B.3 The Total Energy 62
B.4 Dipole Moment 62

C Cusp Conditions for the Jastrow Factor **64**

D Stale Configurations **66**

E Matrix Algebra **68**
E.1 The Local Energy 68
E.2 Simplification 69
E.3 Catch 71

F Improved Sampling **73**

G A Simple SCF Model of CuH **76**

Bibliography **80**

List of Tables

4.1 Orbital exponents ζ for core orbitals 34

4.2 Optimized valence parameters for CuH 39

4.3 Jastrow $b_{i,j}$ parameters 39

4.4 Valence energies and dipole moment 46

4.5 Valence energies and dipole moment—final results 46

5.1 Nonrelativistic estimates of dissociation energy and dipole moment for CuH 53

List of Figures

4.1 CuH valence energy vs. parameter a in the Jastrow factor 36

4.2 Cu valence energy vs. parameter a in the Jastrow factor 36

4.3 Var for uniform sampling 37

4.4 Cu and CuH valence energy vs. c_{ij} on $2p_{\pm 1}$ on H 40

4.5 Cu and CuH valence energy vs. c_{ij} on $4p_0$ on Cu 40

4.6 Cu and CuH valence energy vs. c_{ij} on $4p_{\pm 1}$ on Cu and $2p_{\pm 1}$ on H 41

4.7 Cu and CuH valence energy vs. c_{ij} on $2p_0$ on H 41

4.8 Cu valence energy vs. ζ on $4s$ on Cu 42

4.9 CuH valence energy vs. c_{ij} on $1s$ on H 43

4.10 CuH valence energy vs. c_{ij} on $1s'$ on H 43

4.11 CuH valence energy for optimized parameters 44

4.12 Cu valence energy for optimized parameters 44

5.1 CuH valence energy vs. the 'smoothing' parameter ϵ 50

5.2 Cu valence energy vs. the 'smoothing' parameter ϵ 51

F.1 Original sampling 74

F.2 New sampling 75

G.1 SCF: CuH energy vs. ζ on $4s$ on Cu 77

G.2 SCF: CuH energy vs. ζ on $1s$ on H 78

G.3 SCF: Cu energy vs. ζ on $4s$ on Cu 78

G.4 SCF: Cu energy vs. ζ on $1s$ on H 79

Acknowledgement

I gratefully acknowledge generous support as well as helpful discussions and instructions from my professors S. M. Rothstein and J. Vrbik during my work on this project.

I would also like to thank both of my supervisors and members of Department of Physics for their careful reading of my thesis and for their valuable suggestions.

I appreciate very much help of Prof. Sternin, who introduced me to LaTeX and hence saved me a lot of otherwise more painful wordprocessing.

Also, I would like to thank Mr. Harrison for his support in form of the Thompson-Harrison Graduate Scholarship.

Chapter 1

Introduction

In this work we would like to present a new approach to computation of dissociation energy D_e of CuH using the variational Monte Carlo (VMC) method.

In Chapter 2 we give a historical overview of different *ab initio* methods used in quantum chemistry nowadays. The emphasis is on quantum Monte Carlo (QMC), where VMC belongs to, and one particular 'standard' technique (CAS SCF), which is of special interest for our work, because it provides very accurate theoretical data for CuH to compare our results with[1]. Also, this chapter serves as an introduction for those readers, who are not directly involved in computational chemistry and who may therefore find interesting to learn the context for the presented work.

Chapter 3 describes the basics of Monte Carlo (MC) integration and those aspects of MC techniques, which are essential for VMC. Because our approach to computation of CuH dissociation energy is based on VMC, this chapter represents a technical introduction to the presented work. Also, in this chapter we mention a possible improvement of 'split-τ' technique, which is sometimes used in different QMC algorithms.

The core of our work is described in Chapter 4, where our new '$f \times g$ approach' and its application to D_e of CuH is presented.

Logic of Chapter 4 is as follows:

- First we present the $f \times g$ approach from theoretical point of view, i.e. we introduce

[1] Until September 1994 these were the best theoretical calculations for CuH. However, recently published multiconfiguration second order perturbation theory (see [1]) uses CAS SCF wave function as a reference function and gives even more accurate results for CuH.

basic formulas and give some reasoning and interpretation of these formulas.

- Next we describe actual implementation of the presented technique.
- Finally we present and discuss obtained results.

In Chapter 5 we give some final comments and suggest possible orientation of further research.

And finally, for the sake of completeness, Appendix A describes diffusion QMC (DQMC). Other appendices are intended to give a more detailed insight into some technical problems and their solutions.

Chapter 2

Ab Initio Methods

At the very beginning we should point out that the topic of this chapter is not a component of our research. However, we feel that for some readers, particularly those not involved in quantum chemistry, it may be interesting to learn something more about the position of the presented work among other techniques used in computational quantum chemistry.

First we describe self-consistent-field method, which represents the basic method for most techniques used by computational chemists. Then we proceed to configuration interaction (CI) method. Even though CI is a very powerful approach, its straightforward implementation for many electron systems leads to algorithms with enormous computational requirements. Therefore we mention some modified versions of CI, where the modifications are aimed at increasing efficiency and decreasing computational demands. Some CI based algorithms may be considered 'the main stream' in the present computational chemistry. Others, however, are not so popular and not too many researchers are implementing them.

MC based techniques, described in the last section of this chapter, are among those 'out of the main stream' algorithms. They are, however, capable of competing with CI and even, for few electron systems, superior to it (see [2] and [3]).

2.1 Hartree-Fock Self-Consistent-Field Method

In this section we give an overview of different *ab initio* methods based on Hartree-Fock self-consistent-field (SCF) approach; these methods represent the foundation for computational chemistry nowadays.

In 1928 Douglas Hartree suggested to approximate a molecular wave function ψ as a product of single electron wave functions φ_i chosen in such a way that each electron is experiencing only the average electric field created by other electrons and nuclei. This approach is called Hartree SCF method.

It took two years to prove formally that Hartree's intuitive procedure works. It was found that this procedure effectively minimizes the expectation value of energy

$$\frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} \tag{2.1}$$

for ψ having form of a product of single electron wave functions. The original proof was given by Slater and Fock in 1930. For the proof and a review of the SCF method see e.g. [4].

However, in their paper Slater and Fock also pointed out that the Hartree SCF wave function is not antisymmetric, as required by Pauli exclusion principle, and proposed to write the wave function ψ in the form of an antisymmetrized product (Slater determinant) of spin-orbitals[1] φ_i:

$$\psi = \det(\varphi_i) \stackrel{\text{def}}{=} \frac{1}{\sqrt{n!}} \begin{vmatrix} \varphi_1(x_1) & \varphi_1(x_2) & \dots & \varphi_1(x_n) \\ \varphi_2(x_1) & \varphi_2(x_2) & \dots & \varphi_2(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_n(x_1) & \varphi_n(x_n) & \dots & \varphi_n(x_n) \end{vmatrix} \tag{2.2}$$

[1] By a spin-orbital φ we understand a function of four variables: three components of the position vector $\vec{r}$, and one discrete variable with two possible values α and β describing the spin state of an electron. This four-component variable is in (2.2) denoted as x_i.

Having this form of the wave function ψ we can minimize the expectation value of the energy (2.1) and hence optimize the wave function. To perform the optimization we could derive a set of equations for $\varphi_i(x)$ and then solve them by iterating. Details can be found e.g. in [5, Chapter 13.16].

This approach is referred to as Hartree-Fock SCF method.

Originally, the Hartree-Fock equations were solved by numerical integration and the resulting orbitals were given in form of tables. Also, the calculations were done for atoms only. It was just in 1951 that Roothaan extended the method to molecules and suggested to expand the orbitals using a complete set of basis functions χ_j:

$$\varphi_i = \sum_{j=1}^{\infty} c_{i,j}\, \chi_j$$

Similarly as in Hartree-Fock SCF method, we could derive a set of (Roothaan) equations for the $c_{i,j}$'s, which are, in principle, easy to solve by iterating. Again, details can be found in [5, Chapter 13.16]. In practice only a finite (i.e. incomplete) set of basis functions is used and the term 'SCF wave function' is applied to any wave function obtained by an iterative solution of the Roothaan equations.

Modern SCF calculations consist of following steps:

- choice of basis functions χ_i
- formation of the orbitals $\varphi_i = \sum_j c_{i,j}\, \chi_j$
- formation of configuration (state) function (CSF) $\psi = \det(\varphi_i)$
- optimization of $c_{i,j}$'s by an iterative solution of the Roothaan equations

Even though the Roothaan-Hartree-Fock wave functions are able to give rather good estimates of different molecular properties for small molecules, they become very inaccurate for larger molecules. Typical problems are incorrect values of D_e and improper behavior as the molecule dissociates ($R \to \infty$).

Hence, in order to describe dissociation as well as the structure of larger molecules we have to go beyond the Hartree-Fock approximation and introduce correlation among electrons involving interelectronic coordinates.

2.2 Configuration Interaction in SCF

Perhaps the most common way of improving SCF results is so called configuration interaction (CI) method. Using CI one writes the molecular wave function Ψ as a linear combination

$$\Psi = \sum_i a_i \psi_i \tag{2.3}$$

of the CSFs and uses the variation method to find a_i's.

Although an SCF calculation is rather a routine task nowadays, each CI calculation presents its own special problems. Most of those problems have a common denominator: For a given number of electrons the number of CSFs—the number of terms in (2.3)—increases very fast with increasing the number of the basis functions χ_i. The dependence is approximately described by formula (see [5, Chapter 13.21])

$$\#\text{CSFs} \approx \#\text{basis functions}^{\#\text{electrons}}$$

Therefore, due to physical limits of computers available, full CI calculations[2] are possible only for small molecules and small basis sets. Hence, we have to find a way of reducing the number of CSFs used and/or speeding up convergence of the CI expansion.

Different approaches to solve the mentioned problem have been suggested, e.g.:

- single-double CI (SDCI), which is using only certain CSFs[3] out of all CSFs provided by the chosen basis set, thereby reducing the number of the configuration used

[2] In a full CI calculation we use all configurations provided by the basis functions.

[3] The orbitals used in SDCI are so called singly and doubly excited CSFs. For more details see [5, Section 15.2] and [6, page 255]

- direct configuration interaction method suggested to avoid the problems of dealing with large matrices (see [6, Chapter 7 by B. O. Roos and P. E. M. Seigbahn])
- the graphical unitary group approach (GUGA), see [7, 8, 9], which is speeding up the CI expansion convergence

2.3 Complete Active Space SCF

Another whole branch of techniques aimed to fight the huge number of CSFs in a full CI calculation is based on the possibility to optimize $c_{i,j}$'s and a_i's in the wave function

$$\Psi = \sum_i a_i \psi_i = \sum_i a_i \det(\varphi_i) = \sum_i a_i \det\left(\sum_k c_{i,k} \chi_k\right)$$

simultaneously. That leads to $c_{i,j}$'s optimized for the CI wave function Ψ, rather than to $c_{i,j}$'s optimized for SCF, and hence the CI expansion can converge much faster, thereby allowing substantially fewer CSFs to be included in Ψ.

One of the methods using the idea of simultaneous optimization of $c_{i,j}$'s and a_i's is multiconfiguration SCF (MC SCF).

A commonly used kind of MC SCF is the complete active space SCF (CAS SCF) method (see [10]), which both speeds up the convergence of the CI expansion (as discussed above) and reduces the number of CSFs used in the actual CI calculation.

The reduction of the number of CSFs used is based on the following idea: Divide the orbitals φ_i into inactive and active orbitals. CSFs are then formed in such a way, that all inactive orbitals are doubly occupied in all CSFs[4]. The rest of the orbitals in individual CSFs are taken to be all possible choices from active orbitals (each choice of active orbitals gives one CSF).

[4]Meaning that all active orbitals are used in all CSFs twice, i.e. for both possible spin orientation α and β.

It is reasonable to take the active orbitals as those orbitals that arise from the valence orbitals. For large molecules, however, use of all the valence orbitals as actives gives rise to too many CSFs to be handled and so the number of active orbitals must be limited in such cases.

Abilities of CAS SCF are illustrated e.g. by [11], where the calculation on the C_2 ground electronic state is described. The results are (experimental and Hartree-Fock values given in parenthesis): $R_e/\text{Å} = 1.25(1.24, 1.24), D_e/eV = 6.06(6.3, 0.8), \nu_e/cm^{-1} = 1836\ (1855, 1905), \nu_e x_e/cm^{-1} = 14.9\ (13.4, 12.1)$. Note particularly improvement in D_e, CAS SCF versus SCF.

For CuH, which is of special interest for us, CAS SCF predicts (see [1]) (experimental values given in parenthesis) $R_e/\text{Å} = 1.597\ (1.463^5), D_e/eV = 1.92\ (2.75^6)$.

Also the CAS SCF calculation of HNO—the molecule, which is Hartree-Fock unstable (see [15])—has been successfully carried out (see [16]).

Effectiveness of CAS SCF compare to MC SCF and SCF is noticeable e.g. from [17]: error in D_e of N_2 for SCF is almost 50%, for MC SCF 27% and for CAS SCF 11% (even though the basis set used in MC SCF was much larger than the one used in CAS SCF).

Finally, we can conclude that the popularity of CAS SCF among computational chemists is based on convincing results. However, there are still many problems (e.g. with excited states—see [17], or with very large molecules) which remain to be solved.

2.4 Monte Carlo Methods

The main ideas behind Monte Carlo (MC) techniques were known in the statistics community in the early 1900. However, these ideas were of no practical use until the first powerful computers were made. Since then the MC techniques have become a widely

[5] Reference [12].

[6] Recommended value from [13] according to [14].

employed tool for studying many very different phenomena.

MC methods are very popular especially because of two reasons:

- MC algorithms usually require independent repetition of simple tasks, which is very easy to implement on parallel computers.
- Mathematical formulation of several problems leads to many-dimensional integrals, which are known to be evaluated the most effectively by MC integration.

It is worth noticing that—due to the later reason—MC algorithms, which are in a sense 'random' by their nature, are suitable not only for problems having a probabilistic character, but may be used also for strictly 'deterministic' problems.

History of Quantum MC (QMC) goes as far as year 1949. Then Metropolis [18, 19] reported the first QMC computation. Inspiration for this calculation came from Fermi, who was probably the first to suggest that the similarity between Schrödinger equation and the diffusion equation can be exploited to produce a solution of the Schrödinger equation. However, the results obtained were not any better than those from more traditional numerical methods and the approach faded from the scene for several years.

It was just in 1960s that Kalos developed much more sophisticated MC approach to quantum mechanical problems. Nowadays Kalos' method is called Green's Function MC (GFMC) [20, 21, 22, 23]. Later, in 1970s, Anderson independently solved certain problems in quantum chemistry by using random walks [24, 25, 26, 27]. He was the first who treated fermions by MC. Kalos' and Anderson's algorithms also have an interpretation in terms of Feynman path integrals [28, 29]. The path integral interpretation of the algorithm enables one to reformulate the original approach into a new form called Path Integral QMC [30], which differs from GFMC mainly in the way the kinetic energy operator is treated: In GFMC the kinetic energy operator corresponds to diffusion, whereas

in Path Integral QMC the kinetic energy is represented by a potential between imaginary replica particles.

Yet another version of MC algorithms is variational MC (VMC). This algorithm can be considered as a simplified form of DQMC. Or, alternatively, we can consider VMC to be just a simple MC integration with a bit 'tricky' approach to generating a sample. In this work we describe VMC from the later point of view[7].

We would like to point out that there are QMC algorithms solving the Schrödinger equation for fermions exactly (see [31], [32], [33], [34]). However, for the time being, these algorithms can treat only a few electron problems.

Note that the terminology in this field is not stable yet. Therefore different authors are using terms DQMC, VMC, GFMC with slightly different meanings. Also, sometimes additional adjectives are used to specify fine differences between the algorithms (e.g. 'infinitesimal differential' DQMC in [35], 'quadratic accuracy' DQMC in [36], 'damped-core' quantum MC in [37], 'exact' quantum MC in [31]).

[7]Appendix A describing DQMC also explains VMC as a simplified form of DQMC.

Chapter 3

Variational Monte Carlo

It is relatively well-known fact that—unless a special form of the integrand permits analytical solution—the most effective method for evaluation of multidimensional integrals is the Monte Carlo (MC) method.

Also, evaluation of the variational integral for the total energy of a many particle system with Hamiltonian $\hat{H}$ described by a wave function ψ

$$E = \frac{\int \psi^* \hat{H} \psi \, dV}{\int \psi^* \psi \, dV} \tag{3.1}$$

leads to a multidimensional integral.

Therefore, in this chapter we give an overview of MC integration and discuss a specific approach to a very important point of all MC algorithms—generating of random numbers from a given probability density distribution. We apply the procedure for evaluation of the variational integral (3.1).

3.1 Monte Carlo Integration

MC integration is based on the following formula:

$$\int_M f(x) \, dx = \langle f(x) \rangle \tag{3.2}$$

where $\langle f(x) \rangle$ is the average value of $f(x)$ for x uniformly distributed in the region M.

Therefore, we can estimate the value of the integral on the left side of (3.2) by taking a sample of N points uniformly chosen from M and computing the average value of $f(x)$ for this sample.

It is important to notice that if the main contribution to the integral (3.2) comes from a very small subregion of M where $f(x)$ is large, then the simple MC integration is very inefficient. It is so because most of the points do not come from the 'important' subregion and we spend a lot of time by averaging non significant contributions to (3.2).

The problem described can be solved in the following way: If integrand $f(x)$ has very different values in different parts of the integration region M, do not sample M uniformly. Rather, choose a probability distribution $p(x)$ giving higher probabilities to 'more important' regions of M and sample M according to $p(x)$.

Then, of course, the integral on the left side of (3.2) is not equal to the average value of $f(x)$. Instead, we have to use a modified formula:

$$\int_M f(x)\, p(x)\, dx = \langle f(x) \rangle_{p(x)} \tag{3.3}$$

This approach is known as the importance sampling and can be effectively used to evaluate the variational integral (3.1) written in the form

$$E = \int E_l \frac{\psi^2}{\int \psi^2\, dV}\, dV = \langle E_l \rangle_{\psi^2} \tag{3.4}$$

where

$$E_l = \frac{\hat{H}\psi}{\psi}$$

is a 'local energy', we assumed a real wave function ψ and

$$p = \frac{\psi^2}{\int \psi^2\, dV} \tag{3.5}$$

Note that as ψ approaches the exact solution of the Schrödinger equation, E_l becomes a constant and the importance sampling works very well, because $\langle E_l \rangle_{\psi^2}$ averages similar numbers.

3.2 The Generalized Metropolis Algorithm

In many practical applications, including VMC, we have to deal with the following problem: How does one generate a sample from a discrete[1] probability density distribution p_i ($i \in \{1, 2, \ldots, N\}$)?

If the values of p_i are easy to calculate, the problem is not very difficult: We could, for example, divide the interval $[0, 1]$ into N disjoint line segments with lengths p_i, $i \in \{1, 2, \ldots, N\}$, and generate uniform random numbers between 0 and 1. Each random number would then represent an event i—based on the line segment the random number falls into.

However, in many physically interesting situations it is very difficult to calculate the values of p_i directly. It is much easier to calculate the ratio p_i/p_j for any i and j instead[2].

As originally pointed out in [38], the knowledge of the relative probabilities p_i/p_j for any i and j is sufficient to generate a sample from the probability distribution p_i. It is possible to do so by using the so-called Metropolis algorithm.

In its generalized form the Metropolis algorithm can be described as follows[3]:

- pick an arbitrary initial state i
- being in a state i, propose a move to a state j with a probability $T_{j,i}$
- accept the proposed move from i to j with a probability $B_{j,i}$
- depending on whether the proposed move was accepted or rejected, the new state will be j or i, respectively

[1] Extension to a continuous case is straightforward.

[2] A typical example may be the Boltzman distribution $C \exp[-E/(kT)]$, where E is energy of a system, T is its temperature and k is Boltzman constant. The normalization constant C is usually unknown and very difficult to calculate. However, the constant is not needed if we work only with relative probabilities $\exp[-\Delta E/(kT)]$, where ΔE represents the energy difference between two states.

[3] As implicitly assumed, $0 \leq B_{j,i} \leq 1$ and $T_{j,i}$ is a probability transition matrix, i.e. $0 \leq T_{j,i} \leq 1$ and $\sum_j T_{j,i} = 1$

- repeat the last three steps

It is not very difficult to see that if the detailed balance condition

$$p_i T_{j,i} B_{j,i} = p_j T_{i,j} B_{i,j}$$

is satisfied, then p_i is an equilibrium distribution for the process described[4].

The detailed balance condition can be satisfied by many different choices of $T_{j,i}$ and $B_{j,i}$. A very common choice for $B_{j,i}$ is

$$B_{j,i} = \min\left(1, \frac{p_j}{p_i}\frac{T_{i,j}}{T_{j,i}}\right) \tag{3.6}$$

The selection of the transition probability matrix $T_{j,i}$ varies from problem to problem. In our case we choose the $T_{j,i}$ matrix which makes the acceptance probabilities $B_{j,i}$ as close to one as possible (see the next chapter). In that way we are improving efficiency of the Metropolis algorithm by not allowing configurations to stay 'locked' at one position.

Note that the conditions for $T_{j,i}$ and $B_{j,i}$ which would assure both the stability of the process and its convergence to the probability distribution p_i are not easy to formulate. Throughout this thesis we assume that our choice of $T_{j,i}$ and $B_{j,i}$ gives a stable and convergent algorithm and we do not attempt to prove it.

3.3 Choice of the Transition Probability Matrix

As already mentioned in the previous section, we would like to choose $T_{i,j}$ resulting in the acceptance probabilities $B_{i,j}$ close to one. In the ideal case $B_{i,j} = 1$, and the detailed balance condition becomes

$$p_i\, T_{j,i} = p_j\, T_{i,j}$$

[4]The detailed balance condition means that for the equilibrium distribution p the probability of going from the i^{th} state to the j^{th} state is the same as the probability of going from the j^{th} state to the i^{th} state for all states i and j. Therefore, the probability distribution p is 'balanced'.

Summation of the last equation for all i's gives[5]

$$\sum_i p_i T_{j,i} = p_j$$

or, in the continuous case[6],

$$\int \hat{T}(x,y)\, p(y)\, dy = p(x)$$

In terms of the density ψ^2 (see (3.5)), the requirement on $\hat{T}$ is

$$\int \hat{T}(x,y)\, \psi^2(y)\, dy = \psi^2(x) \tag{3.7}$$

Now, to satisfy the last condition, we do a step, which may look a bit artificial[7], but gives us a hint how to obtain a good $\hat{T}(x,y)$.

Let us consider a partial differential equation

$$\frac{1}{2} \triangle p - \nabla(F(x)p) = \frac{\partial p}{\partial t} \tag{3.8}$$

describing a diffusion process combined with a flow with velocity

$$F(x) = \frac{\nabla\psi(x)}{\psi(x)}$$

It is straightforward to verify that the stationary solution of the equation (3.8) is ψ^2.

The Green function $G(x,y,\tau)$ of the equation (3.8), i.e. a solution of (3.8) satisfying the initial condition

$$G(x,y,0) = \delta(x-y)$$

has the following two properties:

- $\hat{T}(x,y) = G(x,y,\tau)$ satisfies (3.7) for any τ.

[5]Recall that $T_{i,j}$ is a transition probability matrix and therefore $\sum_i T_{i,j} = 1$.

[6]The indices i and j correspond to variables y and x, the matrix $T_{j,i}$ corresponds to an operator $\hat{T}(x,y)$.

[7]In context of DQMC (see Appendix A) this step is very natural.

- Starting from any function[8] p', consecutive application of $\hat{T}(x,y) = G(x,y,\tau)$ (i.e. performing steps in terms of the Metropolis algorithm) changes p' to the stable[9] solution $p = \psi^2$.

Hence, a suitable choice for $T(x,y)$ is $G(x,y,\tau)$ for any τ.

Unfortunately, for a general $F(x)$, i.e. for a general ψ, the analytical form of $G(x,y,\tau)$ is not known. Therefore, we have to find an approximation of $G(x,y,\tau)$ and use this approximation as the transition probability matrix $T_{i,j}$. Then the acceptance probabilities $B_{i,j}$ are not equal to one. However, if the approximation of the Green function $G(x,y,\tau)$ is good enough, the acceptance probabilities are close to one and the Metropolis algorithm is not getting 'locked'.

In a slightly different situation (for DQMC) a formal derivation of an approximation of the Green function $G(x,y,\tau)$, which is accurate up to τ^2, was published in [39, 36, 40]. Another article [41] analyzes different approximations of the Green function (again for DQMC).

However, as explained in [42] these improvements in the Green function approximation have no significant effect on the final simulation because they do not remove singularities of the approximation[10]. Even though [42] suggests how to remove the singularities we decided to use the most simple approximation, which is accurate only up to terms linear in τ.

Here we describe only an intuitive derivation of the approximation. The formal proof can be found e.g. in [43, Chapter Green's Function Monte Carlo Methods].

The equation (3.8) describes two simultaneous processes: diffusion with diffusion

[8] Ignoring some technical details concerning smoothness of p' and its overlap with stationary solutions of (3.8.

[9] We are not going to explore the stability of the solution in any details. We just note that for physically acceptable ψ's the procedure seems to be stable.

[10] Note, that better approximation of the Green function leads to higher acceptance probabilities for any given τ, and hence, by decreasing the serial correlation, it makes the simulation more effective.

coefficient $1/2$ and drift with velocity $F(x)$. Therefore it is natural to expect that a point particle (δ-function as the initial condition) can be described as doing small drift and diffusion steps.

More precisely, if at the beginning the particle is at $x_0 = y$, then after the first drift step it will be at $x_1 = y + F(x_0)\tau$. The consecutive diffusion step can not be described deterministically (because diffusion is a random process) and we can only find a probability distribution for the particle to be somewhere next to the point x_1. This probability distribution is essentially the Green function for the diffusion process in (3.8) itself, i.e.:

$$\frac{1}{(2\pi\tau)^{3N/2}} \exp\left[-\frac{(x-x_1)^2}{2\tau}\right] \tag{3.9}$$

where N is dimension of the space we are moving in (i.e. dimension of vectors $x, y, \ldots$).

Combining together the drift and diffusion steps we come to an approximation of the Green function $G(x, y, \tau)$:

$$G(x, y, \tau) \approx \frac{1}{(2\pi\tau)^{3N/2}} \exp\left[-\frac{(x-y-F(y)\tau)^2}{2\tau}\right] \tag{3.10}$$

According to the previous discussion, this formula represents the transition probability matrix $T_{i,j}$ describing moves in the Metropolis algorithm.

3.4 Algorithm

Based on the previous reasoning, a simple VMC algorithm evaluating the variational integral (3.4) consists of the following steps:

- Generate an ensemble of configurations $\{x_1, x_2, \ldots, x_N\}$ and choose a value of the time step τ.

- For every configuration x_i propose a move to x_i' with the probability

$$G(x_i', x_i, \tau) = \frac{1}{(2\pi\tau)^{3N/2}} \exp\left[-\frac{(x_i' - x_i - F(x_i)\tau)^2}{2\tau}\right]$$

- For every configuration x_i accept the proposed move with the probability

$$\min\left(1, \frac{\psi^2(x_i')}{\psi^2(x_i)} \frac{G(x_i, x_i', \tau)}{G(x_i', x_i, \tau)}\right) \tag{3.11}$$

- Accumulate (ensemble averages of) local energies E_l.

- Repeat the last three steps.

Even though the above described algorithm in principle works as it is, a few comments should be made:

Time Step Choice of the time step τ is critical.

If the time step is very small, the configurations are moving very slowly. This introduces a very large serial correlation for the local energies E_l and decreases efficiency of the simulation.

If the time step is very large, the approximation (3.10) becomes invalid, resulting in very small acceptance probabilities. This again decreases efficiency of the simulation.

There is no general method for choice of the optimal time step. Its value has to be found experimentally. (Section 3.5 describes an approach which makes the simulation more effective by assigning different time steps to different components of x.)

Equilibration Before we start to take the values of the local energy (and other quantities we are interested in) into averages, we have to wait for the ensemble to 'equilibrate'. We usually follow the trend of E_l and start averaging only when it is systematically neither increasing nor decreasing.

Stale Configurations Sometimes, especially at the beginning of the simulation, while the ensemble is not equilibrated properly, some configurations 'refuse' to move (see Section 3.6). Such configurations require a special treatment.

Usually we keep a counter for every configuration, indicating for how many consecutive steps a particular configuration did not move. Then, if the value of the counter reaches a threshold, we let the configuration go (i.e. we accept the proposed move regardless of the acceptance probability). Or, alternatively, we replace the 'stale' configuration by a randomly chosen 'moving' configuration.

This special treatment biases the simulation (it is inconsistent with the Metropolis algorithm described), so we can use the treatment only during the equilibration.

The cause of this problem is discussed in section 3.6.

Errors The Metropolis sampling by its nature introduces some serial correlation, meaning that two consecutive values of E_l (or an other quantity) are not independent[11]. Therefore, in order to properly estimate statistical errors, we have to compute variances based on blocks of local energies rather than of local energies themselves. Size of the blocks is chosen such that their average local energies are likely to be statistically independent.

3.5 Critical Slowdown and Split-τ Technique

Heavy atom systems represent a challenge for all techniques in computational quantum chemistry.

It is possible to show that in standard QMC algorithms CPU time needed to obtain certain accuracy depends on the number of electrons Z in the system as Z^{α}, where estimates for α are between 5.5 and 8 (see [44, 45]).

[11]That becomes obvious e.g. for $\tau \to 0$. Then $x' \approx x$ and therefore also $E_l' \approx E_l$.

Among others, such a steep dependence is due to different spatial scale for the core and valence electrons.

If we choose the time step τ in the algorithm such that the diffusion steps of the innermost electrons are reasonable, the diffusion steps of the outermost electrons are very small and we need many steps to cover the configuration space for the valence electrons properly.

On the other hand, if we choose a time step suitable for the outermost electrons, the innermost electrons are trying to do very long steps. Hence, the formula (3.10) becomes invalid and the simulation becomes inefficient.

The situation is even more complicated due to the drift $F(x)$, values of which vary over several orders of magnitude depending on the position of the electrons relatively to the nodes.

Many authors have been addressing this problem [44]-[53]. Here we mention only the technique of split-τ suggested in [54].

The main idea of the split-τ approach is to divide the electrons into different shells (based on their distance from a nucleus) and to assign different time steps to electrons in different shells.

In terms of the Metropolis algorithm this means a change in the transition probability matrix $T_{i,j}$—without the corresponding change in the matrix $B_{i,j}$ of the acceptance probabilities (see (3.6))! Therefore, the original split-τ algorithm biases the simulation.

Authors of [54] were aware of this problem. However, they knew that the bias is only due to two electrons switching their positions between shells and, as found in [55], the probability of this happening is very small. Therefore, the bias is expected to be negligible, too.

Reassigning the τ's for the electrons immediately after a move has been proposed (i.e. making the corresponding change in the acceptance matrix $B_{i,j}$) removes any possible

bias. However, the reassigning may be very time consuming and therefore we propose a different modification of the original split-τ algorithm.

The modification is based on a simple idea of assigning a value of the time step to individual electrons as a function of their distance from the heavy atom (Cu). If this is done in both the transition probability matrix $T_{i,j}$ and the acceptance probability matrix $B_{i,j}$, the Metropolis algorithm is consistent and generates unbiased ψ^2 distribution.

3.6 Nodal Problems

As discussed in Appendix D, the acceptance probability (3.11) of a move from x to x' approaches 0 for $x \to x_0$, where $\psi(x_0) = 0$. That effectively means that a configuration starting near the nodal surface of ψ is 'locked', i.e. it 'refuses' to move. Such a configuration becomes a 'stale' configuration (see [56]).

However, once the ensemble of configurations is equilibrated, almost all configurations are further from the nodal region of ψ and the problem vanishes.

Unfortunately, in the case of many electron systems the nodal surface of ψ has a very complicated structure and the problem with the stale configurations may reoccur even after the equilibration. These, more often than not, give rise to 'crazy' values of the local energy. For the time being we do not know how to handle this problem correctly. We just truncate those few values of the local energy, which seem to be 'outlayers'.

3.6.1 Variable-τ Algorithm

We tried to solve problem of the 'locked' configurations by a variable-τ algorithm, which would adjust the time step τ according to the actual position x[12].

The most straightforward idea is to keep the drift and diffusion parts of the step for

[12]For another approach see Appendix F.

each electron approximately the same[13], i.e.

$$\sqrt{\tau} \approx F\,\tau$$

for each electron[14], where $\sqrt{\tau}$ is a typical diffusion step (see (3.9)).

This leads to

$$\tau \approx \frac{1}{F^2}$$

or, to avoid problems from the small values of F,

$$\tau \approx \frac{1}{F^2 + constant}$$

However, we discovered that this simple choice does not solve our problem. It is true, that the electrons do not get 'locked', but only from a 'local point of view.' The electrons keep moving, but in a vicinity of the nodal surface τ becomes so small that the electrons are *effectively* 'locked.'

Because we were not able to come with an idea solving this 'local versus global' problem, we left the problem open.

[13]It is natural to expect that for this choice the simultaneous simulation of the drift and diffusion is the most effective.

[14]Note that this implicitly means a version of split-τ algorithm, because components of F are not the same for all the electrons.

Chapter 4

$f \times g$ Approach to D_e

In this chapter we will pay attention to the non-relativistic, time independent electronic Schrödinger equation

$$\hat{H}\psi = E\psi$$

where for N_e electrons and N_n nuclei with charges Z_A the Hamiltonian is[1]

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N_e}\triangle_i - \sum_{i=1}^{N_e}\sum_{A=1}^{N_n}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N_e}\sum_{j>i}^{N_e}\frac{1}{r_{ij}} + \sum_{A=1}^{N_n}\sum_{B>A}^{N_n}\frac{Z_A Z_B}{r_{AB}}$$

More precisely, we describe a method for evaluation of the variational integral for the total energy (3.4).

4.1 Inspiration

Our objective is to arrive at a technique which provides an accurate value for the dissociation energy (D_e) for a molecule containing a heavy atom. The challenge here is that D_e is typically of order of units while the total energy is of order of thousands of units.

Here we describe one such scheme which rests on the separation of the energy due to the core electrons and the valence ones[2]. Furthermore, the energy due to the core

[1] Note that we are using atomic units with the mass of electron $m_e = 1$, the charge of electron $e = 1$, and $\hbar = 1$.

[2] According to quantum mechanics all electrons are identical and indistinguishable and we can not talk about the 'core' or 'valence' electrons. However, in the course of simulation as described in the previous chapter individual electrons have some identity, can be distinguished and their 'movement' (though not physical) can be tracked. In this sense we can talk about the core and valence electrons. Question whether it is or it is not possible to give some deeper meaning to notion of the 'valence' and 'core' electrons, however interesting it may be, is not discussed in this work.

electrons (the bulk of the total energy) cancels when taking the energy difference (e.g. between Cu and CuH) in order to get D_e.

Another important feature of the presented scheme is that it was intended to be computationally effective in the following sense: We believe that the core electrons are not important for the dissociation energy, which is mainly determined by the valence electrons. Therefore, it seems to be reasonable not spend a lot of time by dealing with the core electrons and to pay much more attention to the valence ones. At the same time it would be very good if we could benefit from the core electrons not moving very often by having an effective way of calculating moves of the valence electrons.

Indeed, as suggested by [57], it is possible to devise a wave function approximation having the two properties (the separability of the core and valence energy, and the possibility to calculate the moves of the valence electrons faster using the fact that the core electrons are not moved very often[3]). The suggestion is

$$\psi(x,y) = f(x)\frac{g(x,y)}{\sqrt{\int g^2(x,y)\,dy}} \tag{4.1}$$

where $f(x)$ is the Slater determinant for the core electrons and $g(x,y)$ is the Slater determinant (modified by a Jastrow correlation function) for both the core and valence electrons (denoted x and y, respectively).

This Ansatz also uses the assumption that the term

$$\frac{f(x)}{\sqrt{\int g^2(x,y)\,dy}}$$

is close to 1. Then, the approximation is essentially equivalent to $\psi(x,y) = g(x,y)$ (the Slater determinant and Jastrow factor)—a very common starting point for many methods in quantum chemistry (see the discussion about the *ab initio* techniques in Chapter 2).

[3]For a detailed explanation of how the effectiveness is achieved and why it does not work so well as expected see Appendix E.

In addition to the presented original reasoning we would like to show another way of obtaining (4.1), which perhaps provides deeper insight into the approach.

Let $\psi(x,y)$ be the exact solution of the Schrödinger equation $\hat{H}\psi = E\psi$, where x and y denote the core and valence electrons respectively. Then $\psi^2(x,y)$ is a probability density[4] of observing the core electrons at x and the valence electrons at y.

If the core and valence electrons were independent, then

$$\psi^2(x,y) = f^2(x)\, h^2(y) \tag{4.2}$$

where $f^2(x)$ and $h^2(y)$ are probability densities of observing the core and valence electrons at x and y respectively. However, the electrons are not independent and equation (4.2) holds only approximately.

Assuming that

- the core electrons are almost not influenced by the valence electrons
- the valence electrons are influenced by the core electrons in a nonnegligible way

we can improve the approximation (4.2) simply by taking $h^2 = h^2(y|x)$, where $h^2(y|x)$ is a conditional probability distribution function, i.e. we explicitly express the fact that the probability density of observing the valence electrons at x is dependent on the core electron positions y.

Instead of blindly trying different $f^2(x)$ and $h^2(y|x)$ we can use information which is already available: $f^2(x)$ equal to the Slater determinant[5] squared, and $h^2(y|x)$ equal to the Slater determinant[6] squared and normalized such that $\int h^2(x,y)\, dy = 1$.

Hence, we come to a suggestion

$$\psi^2(x,y) = f^2(x)\frac{g^2(x,y)}{\int g^2(x,y)\, dy} \tag{4.3}$$

[4]Assuming the real wave function $\psi(x,y)$.

[5]See the discussion about the *ab initio* techniques in Chapter 2.

[6]Eventually modified by a Jastrow factor to account for correlation effects.

where

$$h^2(y|x) = \frac{g^2(x,y)}{\int g^2(x,y)\,dy}$$

is already written in the normalized form, what is essentially the original suggestion (4.1).

4.2 Basic Formulas

The previous reasoning lead us to the approximation of the wave function $\psi(x,y)$

$$\psi(x,y) = f(x)\frac{g(x,y)}{\sqrt{\int g^2(x,y)\,dy}} \tag{4.4}$$

After some tedious but straightforward algebra the expression for the total energy E

$$E = \frac{\int \psi^* \hat{H} \psi \, dV}{\int \psi^* \psi \, dV}$$

becomes (see Appendix B)

$$E = \left\langle -\frac{1}{2}\frac{\triangle_x f(x)}{f(x)} + V_c \right\rangle_c + \left\langle -\frac{1}{2}\frac{\triangle_x g(x,y)}{g(x,y)} + V_v \right\rangle_{v,c} + \frac{1}{2}\langle \mathrm{Var} \rangle_c + V_{nn} \tag{4.5}$$

where

$$\mathrm{Var} = \left\langle \left(\frac{\nabla_x g(x,y)}{g(x,y)} \right)^2 \right\rangle_v - \left\langle \frac{\nabla_x g(x,y)}{g(x,y)} \right\rangle_v^2 \tag{4.6}$$

and $\langle q \rangle_c$ means average value of quantity q over all core configurations, i.e.

$$\langle q \rangle_c = \frac{\int q\, f^2(x)\,dx}{\int f^2(x)\,dx} \tag{4.7}$$

Similarly $\langle q \rangle_v$ refers to the average value of q over all valence configurations (conditional upon a core configuration), i.e.

$$\langle q \rangle_v = \frac{\int q\, g^2(x,y)\,dy}{\int g^2(x,y)\,dy} \tag{4.8}$$

The term V_c refers to the potential energy contributions containing only interactions between the core electrons and nuclei and and among the core electrons. Similarly,

V_v denotes those potential energy contributions which include an interaction with the valence electrons, i.e. the valence-valence, valence-core and valence-nuclei interactions.

Interpretation of the individual terms in the equation (4.5) is as follows:

- The first two terms represent the average kinetic and potential energy of the core electrons. It is important that these terms 'do not know' anything about the valence electrons. As we will see later, this separation of the core energy has very useful implications. Therefore it was also an essential guiding idea for choice of the approximation (4.4).

- The second two terms represent the average kinetic and potential energy of the valence electrons. These two terms 'know' about the core electrons (and the core wave function $f(x)$) via the core-valence interactions in V_v and the averaging $\langle \ldots \rangle_c$.

- As follows from the derivation of (4.5) and (4.6) in Appendix B, the third term represents a correction to the core kinetic energy due to the valence electrons. From now on we will refer to this term as the variance term[7]. The correction arises because of the core-valence interaction in $g(x,y)$. If $g = g(y)$, i.e. the valence electrons would be independent of the core ones, there would be no such term.

- Finally, the last term is nuclear-nuclear repulsion. This term depends only on the geometry of the molecule and not on the wave function. For CuH $V_{nn} = 29/R$, where R is the distance between Cu and H nuclei. For Cu $V_{nn} = 0$ (the term is not present).

[7]Reason for the name 'variance term' is obvious: If one defines the gradient

$$F(x,y) = \frac{\nabla_x g(x,y)}{g(x,y)}$$

then the 'variance term' is the variance of F over all valence configurations (averaged over all core configurations and divided by two).

4.3 Separation of the Valence and Core Energy

Note that in (4.5) the first two terms do not depend on any parameters in $g(x, y)$. These terms depend only on the function $f(x)$, which describes the core electrons. Therefore it is natural to call the first two terms in (4.5) the core kinetic and core potential energy respectively. Similarly, we say that the second two terms represent the valence kinetic and valence potential energy.

Hence, using the terminology introduced, the total energy corresponding to the wave function (4.4) consists of four parts: the (kinetic and potential) core energy, the (kinetic and potential) valence energy, the variance term and the nuclear-nuclear repulsion.

The fact that the core energy is independent of the valence parameters[8] has important consequences:

- It enables us to optimize the valence part of the wave function ψ without a need to compute the total energy[9].

  This aspect is very important, because the valence parameters have only a very small influence (in order of 0.1 a.u.) on the total energy (in order of 1600 a.u.)[10]. Therefore, in order to achieve a precision needed to recognize these very small changes of the total energy, one would need a lot of CPU time and the whole optimization process would become very inefficient.

- The separation of the core energy may become a very effective tool for the computation of dissociation energy.

  In our case the Cu core is described by the same function $f(x)$ as the CuH core

[8] Parameters involved in $g(x, y)$, which describes the valence electrons.

[9] Assuming that the core part $f(x)$ of the wave function has been optimized sufficiently. However, one can hope that this part is not very critical and that it is possible to describe the chemically important valence electrons using only a rough approximation for the core.

[10] The actual values are for CuH.

(corresponding functions $g(x, y)$ are, of course, different), which leads to the exact cancellation of the CuH and Cu core energies.

Unfortunately, as we will further see, we have to pay for these advantages by having difficulties with the variance term.

4.4 Philosophy of the Program

After choosing the approximation of the wave function (4.4) we could evaluate the corresponding energy (3.4) by directly applying the technique described in Chapter 3, i.e. by diffusion and drift[11] of both the core and valence electrons at the same time. However, this approach would not give us the core and valence energy separation with advantages of the exact cancellation of the core energy. Also, we could not exploit the special form of ψ to get the efficiency improvement described.

Therefore, we turn our attention to an alternative expression for the energy (4.5). This expression suggests not to calculate the energy as a single average (integral over all possible configurations of both the core and valence electrons) but as a double average (average separately over the core and valence electrons).

Indeed, this is the underlying idea in our program:

- First we generate an ensemble of core configurations (x's) distributed according to the (normalized) probability distribution function $f^2(x)$ which enables us to evaluate integrals of the form (4.7) in (4.5). We do so by using the technique described in Chapter 3 for $p = f^2$.

[11] Using $\nabla\psi/\psi$.

- Then, for each core configuration x we generate corresponding valence configurations y's according to the conditional probability distribution function

$$p(y|x) = \frac{g^2(x,y)}{\int g^2(x,y)\,dy}$$

  Again, we do so by using the technique described in Chapter 3 for $p = p(y|x)$. This enables us to evaluate integrals of the form (4.8) in (4.5).

- We repeat the above described two steps to obtain acceptable statistical errors of the energies.

4.5 The Variance Term

4.5.1 Origin of the Variance Term

If we did not have the variance term, the valence energy concept, as introduced here, would be very similar to the one used in an effective Hamiltonian approach: We would have the valence electrons in a potential representing the overall influence of the core electrons.

In our case the potential would not be given directly in a functional form. Rather, it would be given indirectly via the functions $f(x)$ and $g(x,y)$, which are describing the core electrons and their influence on the valence electrons.

However, there is no obvious reason for an additional (variance) term in the expression for the (valence) energy and we have to accept existence of this term as a consequence of the approximation (4.4). As follows from the derivation of (4.5) and (4.6) in Appendix B, the variance term represents a correction to the core kinetic energy due to the valence electrons.

4.5.2 Evaluation of the Variance Term

First recall the form of the first part of the variance term (4.6):

$$\mathrm{Var} = \left\langle \left(\frac{\nabla_x g(x,y)}{g(x,y)} \right)^2 \right\rangle_v - \cdots$$

where for any function q

$$\langle q \rangle_v = \frac{\int q\, g^2(x,y)\, dy}{\int g^2(x,y)\, dy} \tag{4.9}$$

and therefore (only the 'critical' term is written explicitly)

$$\mathrm{Var} = \frac{\int (\nabla_x g(x,y))^2\, dy}{\cdots} + \cdots$$

At this point it is very important to realize that the integrand in the last expression is nonzero for (x,y) at the nodes of $g(x,y)$ and zero for (x,y) at 'the middle'[12] of the nodal regions $g(x,y)$.

Therefore the importance sampling according to ψ^2 which was working efficiently for computation of the kinetic and potential energy[13] is not efficient any more.

An easy way to avoid this problem is to replace the importance sampling (4.9) by a uniform sampling scheme

$$\langle q \rangle_v = \frac{\langle q g^2(x,y) \rangle}{\langle g^2(x,y) \rangle} \tag{4.10}$$

while evaluating the variance term.

Finally a note for those familiar with DQMC (see Appendix A): It is important to realize that evaluation of the variance term using DQMC is going to be a very difficult task because DQMC replaces (4.9) by

$$\langle q \rangle_v \approx \frac{\int q\, g_\tau^2(x,y)\, dy}{\int g_\tau^2(x,y)\, dy} \tag{4.11}$$

[12]By 'the middle' of a nodal region we understand a point, where $g(x,y)$ reaches its local extreme (maximum or minimum), i.e. $\nabla_x g(x,y)$ is zero.

[13]Recall that both the kinetic energy density $\psi^* \hat{T} \psi$ and the potential energy density $\psi^* \hat{V} \psi$ are zero at the nodes of ψ.

where g_τ^2 is a τ-biased g^2 distribution. Then extrapolation to zero τ is made.

However, the variance term involves integral

$$\int \left(\frac{\nabla_x g(x,y)}{g(x,y)} \right)^2 g_\tau^2(x,y) dy$$

the value of which is—due to the distribution g_τ^2 not having any nodes and g^2 having some nodes—infinity!

A similar problem has to be solved also in 'traditional' DQMC (without the variance term). However, the singularities encountered there are of the form $1/g$ and can be removed by a properly devised τ-dependent truncation scheme. For the variance term the singularity is $1/g^2$ and it is much more difficult to deal with it.

4.6 The Exact Solution

After introducing the approximation (4.4) it is very natural to ask ourselves: 'Can we still reproduce the exact solution of the Schrödinger equation by our approximate formula? Or did we restrict the wave functions we can simulate in such a way that we can no longer hope to get the exact solution?'

By squaring equation (4.4) and consecutive integration over y we can easily see that in order to get the exact solution of the Schrödinger equation we have to satisfy

$$f^2(x) = \int \psi_{exact}^2(x,y)\, dy$$

Therefore, in principle, without optimizing $f(x)$ we can not hope to approximate the exact solution of the Schrödinger equation with an arbitrary precision.

And to answer the original question, it is easy to see, that the choice of

$$f(x) = \sqrt{\int \psi_{exact}^2(x,y)\, dy}$$

and

$$g(x,y) = \psi_{exact}(x,y)$$

leads to $\psi = \psi_{exact}$.

Hence, we can say that our choice of the functional form for ψ is—in principle—not restricting a possibility to get the exact solution. However, it is not possible to get the exact solution by optimizing $g(x, y)$ only.

4.7 Choice of the Function f

As the function f we decided to use product of the two Slater determinants[14] of 14 Cu-centered Slater-type orbitals[15] (STOs) (see [58]) $1s, 2s, 2p_{\pm1,0}, 3s$, $3p_{\pm1,0}$, and $3d_{\pm2,\pm1,0}$ with the normalized form

$$Nr^{n-1}Y_l^m(\theta,\phi)e^{-\zeta r}$$

where n, l, m are quantum numbers, $Y_l^m(\theta,\phi)$ are real normalized spherical harmonics

$$Y_l^m(\theta,\phi) = \sqrt{\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}} P_l^{|m|}(\cos(\theta)) \begin{cases} \sqrt{2}\sin(|m|\phi) & m < 0 \\ 1 & m = 0 \\ \sqrt{2}\cos(|m|\phi) & m > 0 \end{cases}$$

P_l^m are associated Legendre functions

$$P_l^m(x) = \frac{1}{2^l l!}(1-x^2)^{m/2}\frac{d^{l+m}}{dx^{l+m}}(x^2-1)^l$$

and N is the normalization factor

$$N = \frac{(2\zeta)^{n+\frac{1}{2}}}{\sqrt{(2n)!}}$$

The orbital exponents ζ were those according to [59] and are listed in table 4.1.

[14] One for spin up and one for spin down electrons. As demonstrated in [5, Chapter 10.7] for purpose of the energy evaluation the single Slater determinant of spin-orbitals (2.2) leads to product of two Slater determinants of spatial orbitals for the two spin orientations.

[15] Except for the radial part, these orbitals are the same as hydrogen like orbitals.

orbital	ζ
$1s$	28.3290700
$2s$	10.5334040
$2p_{\pm 1}$	12.5540500
$2p_0$	12.5541300
$3s$	5.1563900
$3p_{\pm 1}$	4.8759360
$3p_0$	4.8758920
$3d_{\pm 2}$	4.2017730
$3d_{\pm 1}$	4.2014738
$3d_0$	4.2014249

Table 4.1: Orbital exponents ζ for core orbitals

4.8 Choice of the Function g

As for the function g, we followed approach of [60] and used product of so called Jastrow factor (see [61]) and the two Slater determinants of 15 orbitals: 14 STOs as for f and one valence orbital given as a linear combination of Cu-centered $4s$ STO and H-centered $1s, 1s', 2p_0$ STOs[16].

We decided to use experimental geometry, i.e. internuclear distance $R = 2.764$ a.u. (see [62])[17].

Note, that we used H-centered orbitals also for Cu. We decided to do so in order to avoid a possible basis-set superposition error (see [5, Chapter 17.1]). This error arises from the fact that—if we did not use the H-centered orbitals also for Cu—higher number of the orbitals (and hence the variational parameters) in CuH relatively to Cu could artificially lower the energy of CuH compared to the energy of Cu. For more discussion about the basis-set superposition error and its correction see [63].

[16] In Appendix G we demonstrated that the Cu centered $4s$ orbital and H centered $1s$ orbital are sufficient to give a bond for the CuH molecule.

[17] We also tried to optimize the internuclear distance. However, we were not able to obtain any optimal value.

However, as suggested by our results (see Chapter 4.11), H-centered orbitals on Cu play in our case only very little (if any) role. This may be due to symmetry properties of Cu: As a spherically symmetric atom, Cu can not benefit from any orbitals disrupting this symmetry.

The Jastrow factor, which allows explicitly for electron-electron correlation, is given by

$$\exp\left(\sum_{i<j} \frac{a_{i,j}\, r_{i,j}}{1 + b_{i,j}\, r_{i,j}}\right)$$

where the sum is over all possible pairs of electrons i, j, $r_{i,j}$ is distance of the i^{th} and j^{th} electrons and $b_{i,j}$'s are variational parameters.

Parameters $a_{i,j}$ may be also treated as variational parameters. However, some other researchers (see e.g. [42]), prefer to take

$$a_{i,j} = \begin{cases} 1/4 & \text{if } i \text{ and } j \text{ are of like spin} \\ 1/2 & \text{if } i \text{ and } j \text{ are of unlike spin} \end{cases}$$

This choice of the constants $a_{i,j}$ is determined by the requirement that the wave function have the correct cusp conditions as the two electrons approach each other, i.e. the singularity in the local potential energy is canceled by a term in the local kinetic energy so that the (total) local energy is nonsingular (see Appendix C).

In our simulations we let $a_{i,j} = a/4$ for the spin-like electrons and $a_{i,j} = a/2$ for the spin-unlike electrons where a was treated as one of the variational parameters. However, as suggested by Fig. 4.1 and 4.2, it is difficult to find values of a better than the theoretical value $a = 1$.

4.9 Estimate of the Variance Term

As already predicted in Section 4.5.2, the attempt to estimate the variance term using the importance sampling introduced for energy computation fails because of very high

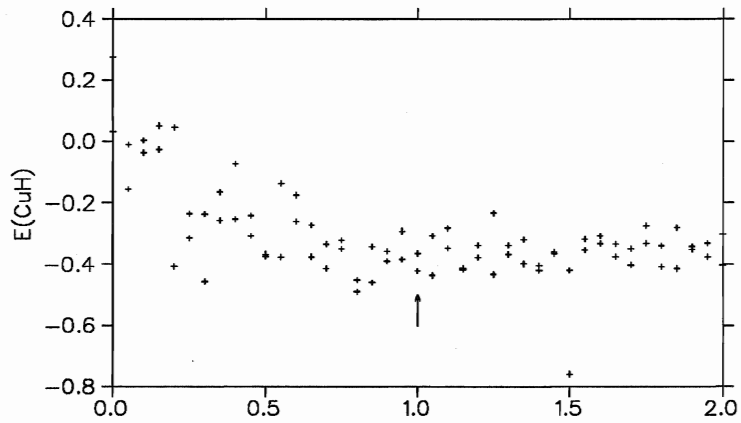

Figure 4.1: CuH valence energy (a.u.) vs. parameter a in the Jastrow factor for optimized parameters (every point represents average of 10 blocks; see Chapter 4.11 for more details)

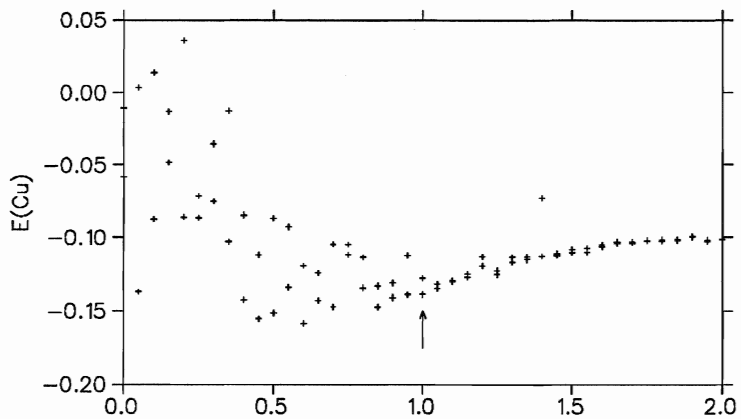

Figure 4.2: Cu valence energy (a.u.) vs. parameter a in the Jastrow factor for optimized parameters (every point represents average of 10 blocks; see Chapter 4.11 for more details)

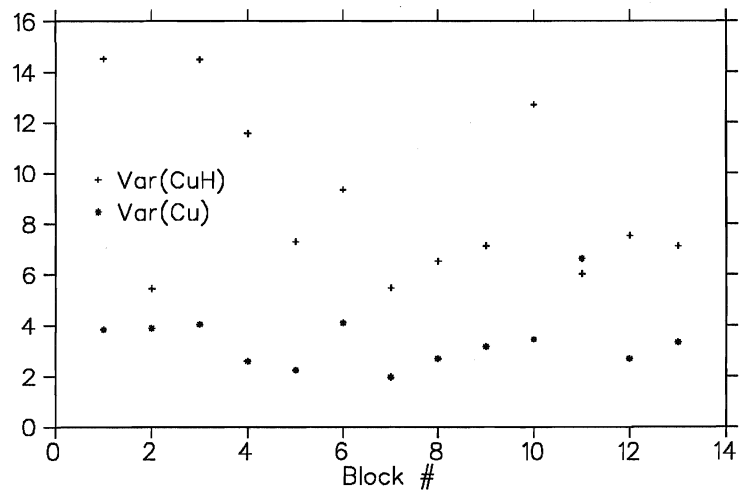

Figure 4.3: Var (a.u.) for uniform sampling for optimized parameters (every block corresponds to average Var for an ensemble of 100 configurations; individual values of Var were calculated from 1000 electron configurations; position of every electron was chosen randomly from a cube with side of 10 a.u. and center at Cu nucleus)

statistical errors.

On the other hand, as we can see from data in Figure 4.3 obtained for optimized parameters (see below), the estimate of the variance term using uniform sampling (4.10) leads to rather consistent results.

4.10 Optimization

In the previous discussion of the variance term we already mentioned that this term is in some sense 'surprising'. Attempts to estimate the value of the variance term also lead to unexpected results: The value found was much larger than the valence (potential and kinetic) energy and also much larger than chemically expected energies involved in the valence shell. Also the difference of the variance terms for Cu and CuH gives the dissociation energy wrong by order of magnitude.

For these reasons we decided to exclude the variance term from any other considerations, and reflect on how to deal with it later.

After excluding the variance term from the expression for the total energy E in (4.5) the only valence-parameter dependent contribution is given by the valence energy. Therefore, in order to optimize the wave function (4.4), we have to minimize the valence energy.

To do so, we decided to use very inefficient but robust 'graphical approach': We computed the valence energy for several values of a chosen parameter and then estimated the optimal value of the chosen parameter from a graph of valence energy vs. the parameter.

The main reason for using this approach was lack of any initial knowledge about the shape of the valence energy surface (i.e. which parameters are more and which parameters are less important, and what are sensible values of different parameters). And, according to a common sense and also our previous experience, in such a situation it is better to start with a robust technique.

4.11 Results

The final optimized values of the valence parameters for CuH are summarized in table 4.2. The optimized values for Cu are discussed below.

The final optimized values of Jastrow $b_{i,j}$ parameters are in table 4.3

Figures 4.4-4.10 depict the Cu and CuH valence energy dependence on some valence parameters as obtained in the final stages of the optimization. Note, that to generate each of these graphs took approximately 24 hours of CPU time on RISC based microprocessor R4400 running at 150 MHz and equipped with floating point unit R4010. To optimize all (valence) parameters took approximately one month of CPU time.

Typically, every point in these figures we obtained is an average of 5 blocks, each block consisting of 100 iterations of the valence electrons. The energies from the first 20

	Cu centered			H centered			
orbital	$4s$	$4p_{\pm 1}$	$4p_0$	$1s$	$2p_{\pm 1}$	$2p_0$	$1s'$
ζ	2.01	X	X	1.14	X	1.00	0.81
$c_{i,j}$	1.00	0.00	0.00	1.60	0.00	0.25	0.90

Table 4.2: Optimized valence parameters for CuH

$b_{i,j}$: valence vs.	valence	1^{st} shell	2^{nd} shell	3^{rd} shell
CuH	2.0	3.0	3.5	1.5
Cu		3.0	3.0	3.5

Table 4.3: Jastrow $b_{i,j}$ parameters

iterations were ignored. These iterations were intended for equilibration. Between every 2 blocks we did 10 iterations of the core electrons. This was done for an ensemble of 100 configurations.

The electrons were moved using the following values of the time step τ: 0.0004 for $1s$ electrons, 0.004 for $2s$ and $2p$ electrons, 0.024 for $3s$ and $3p$ electrons, 0.036 for $3d$ electrons and 0.4 for the valence electrons (all times are in atomic units).

We found that none of the four H-centered orbitals on Cu ($1s$, $2p_{\pm 1}$, $2p_0$ and $1s'$) was able to improve the Cu valence energy. Similarly, Cu-centered $4p$ orbitals turned out to have no observable contribution to the Cu energy. Hence, in addition to the Jastrow parameters, the only valence parameter for Cu was ζ on the Cu-centered $4s$ orbital. The optimal value for this parameter was found to be 1.2 (see Fig. 4.8).

For CuH we found that the valence energy is very sensitive on the parameters ζ on Cu-centered $4s$, H-centered $1s$ and $1s'$ and therefore these parameters were easy to optimize. However, the CuH valence energy was not very sensitive on ζ on H-centered $2p_0$ orbital. Also, the optimization of c_{ij} on $1s$ and $1s'$ orbital on H turned out to be

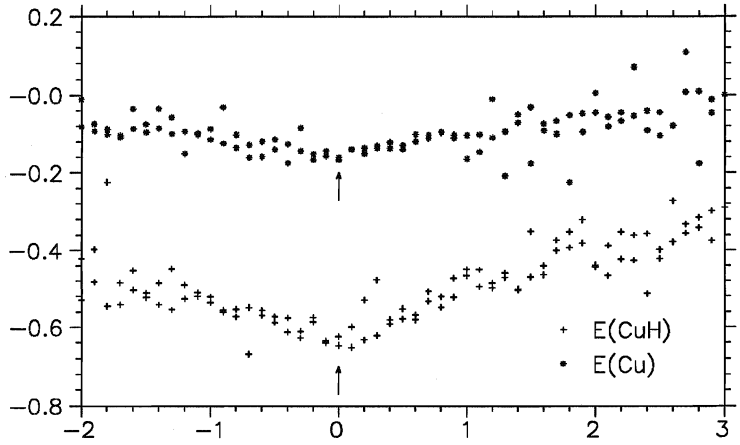

Figure 4.4: Cu and CuH valence energy (a.u.) vs. c_{ij} on $2p_{\pm 1}$ on H (see Chapter 4.11 for more details)

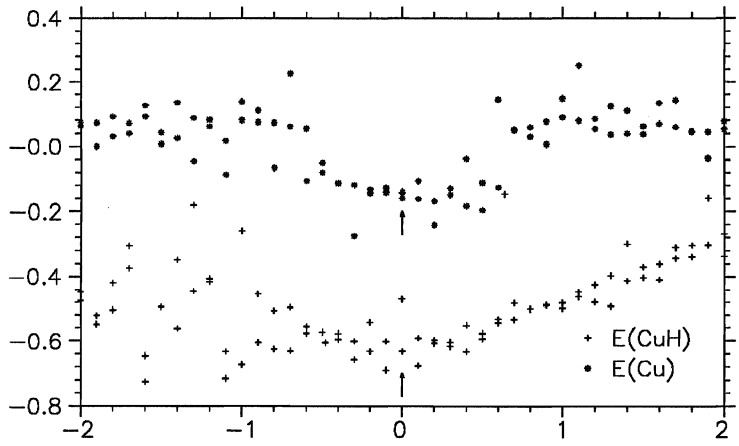

Figure 4.5: Cu and CuH valence energy (a.u.) vs. c_{ij} on $4p_0$ on Cu (see Chapter 4.11 for more details)

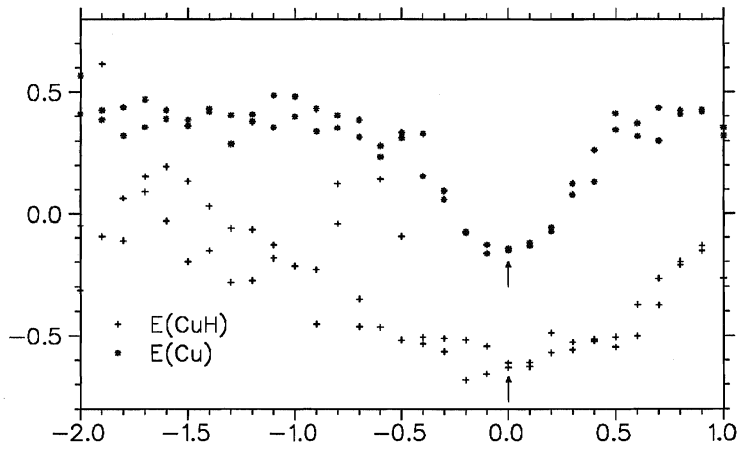

Figure 4.6: Cu and CuH valence energy (a.u.) vs. c_{ij} on $4p_{\pm 1}$ on Cu and $2p_{\pm 1}$ on H (see Chapter 4.11 for more details)

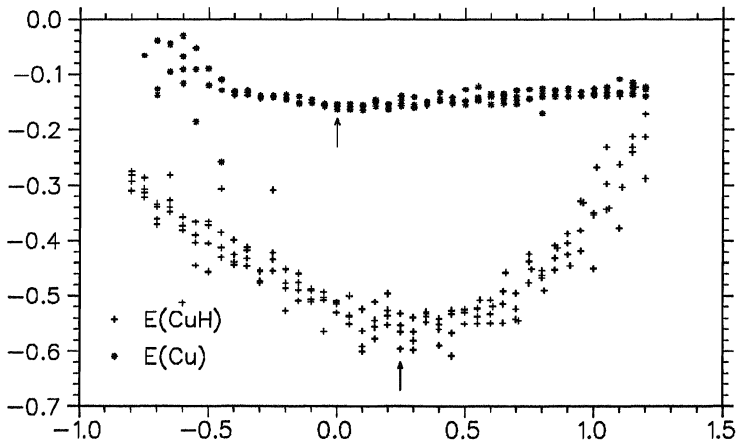

Figure 4.7: Cu and CuH valence energy (a.u.) vs. c_{ij} on $2p_0$ on H (see Chapter 4.11 for more details)

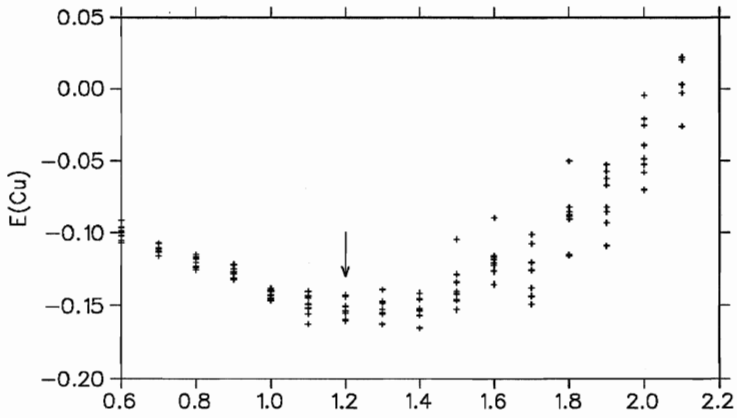

Figure 4.8: Cu valence energy (a.u.) vs. ζ on $4s$ on Cu (see Chapter 4.11 for more details)

rather difficult (see Fig. 4.9 and 4.10). As indicated by Fig. 4.7, the CuH valence energy is sensitive on the c_{ij} on H-centered $2p_0$ orbital and this parameter was easy to optimize.

Once we obtained the optimal parameters we did a long run to calculate Cu and CuH energies with sufficiently small errors. Results are ploted in Fig. 4.11 and 4.12. The energies are $E_{\mathrm{CuH}} = -0.624 \pm 0.006$ a.u. and $E_{\mathrm{Cu}} = -0.164 \pm 0.003$ a.u.

Taking into account the exact energy of the hydrogen atom, $E_{\mathrm{H}} = -0.5$ a.u., we get the dissociation energy $D_e = E_{\mathrm{Cu}} + E_{\mathrm{H}} - E_{\mathrm{CuH}} = -0.040$ a.u. with standard error 0.009 a.u. Unfortunately, this value is very different (with even incorrect sign predicting spontaneous dissociation of CuH into Cu and H) from the experimental value 0.101 a.u. (see [13]).

Value of the dipole moment from the final run, 1.08 ± 0.03 a.u., is in better agreement with the best CI calculations (see e.g. [64]) giving the value 1.2 a.u.

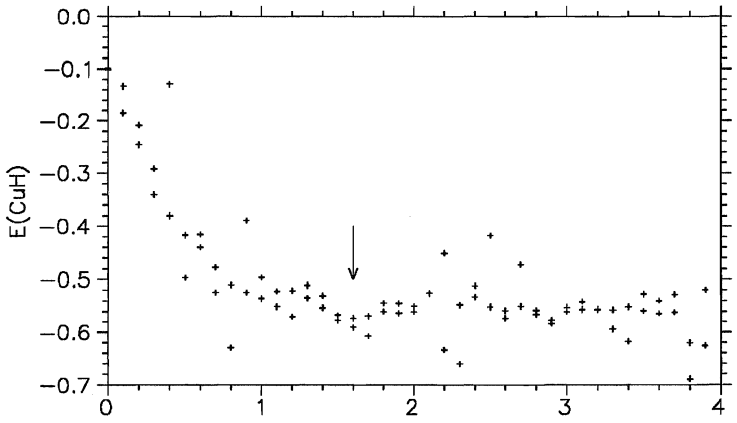

Figure 4.9: CuH valence energy (a.u.) vs. $c_{i,j}$ on $1s$ on H (see Chapter 4.11 for more details)

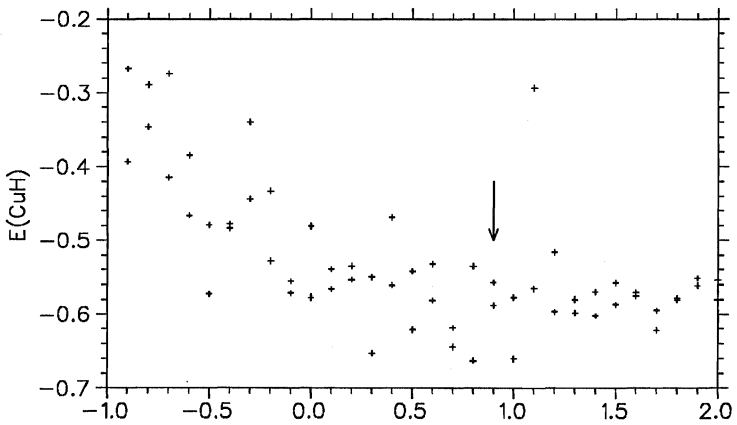

Figure 4.10: CuH valence energy (a.u.) vs. $c_{i,j}$ on $1s'$ on H (see Chapter 4.11 for more details)

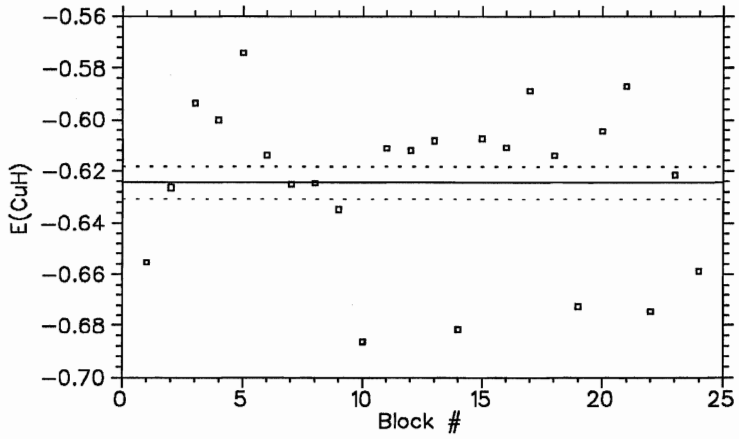

Figure 4.11: CuH valence energy (a.u.) for optimized parameters (solid line represents the average and dashed lines its standard error; see Chapter 4.11 for more details)

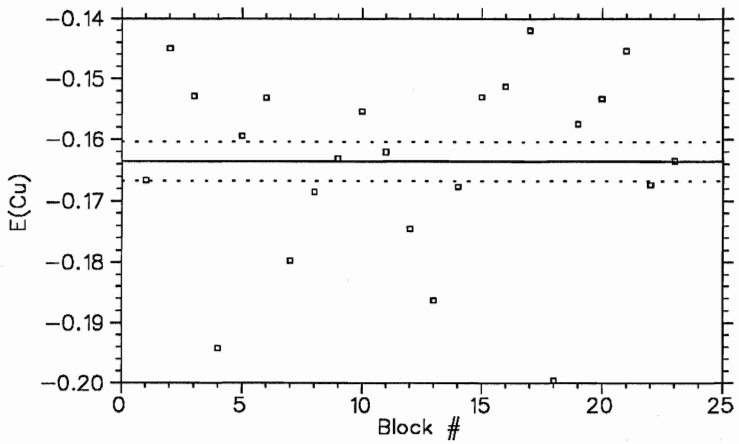

Figure 4.12: Cu valence energy (a.u.) for optimized parameters (solid line represents the average and dashed lines its standard error; see Chapter 4.11 for more details)

4.12 Results—Improved Sampling

As already mentioned in Section 3.6, the original Metropolis algorithm does not handle the nodal region of the wave function properly (see also Appendix F). To find out how critical this problem is we decided to implement the modification of the sampling scheme according to [42] and to calculate the valence energies and dipole moment for the optimized parameters using this improved sampling.

The results of 6 runs, 300 blocks[18] each, are presented in table 4.4. These data confirm that the original sampling does not work properly and leads to a biased estimate of the valence energies: Improved sampling gives the valence energies $E_{\mathrm{CuH}} = -0.898 \pm 0.012$ a.u. and $E_{\mathrm{Cu}} = -0.249 \pm 0.004$ a.u., i.e. both energies decreased significantly. However, the value of the dipole moment changed only very slightly (to 1.045 ± 0.015 a.u.).

Seeing the bias of the original sampling we decided to reoptimize the valence parameters. However, we found only a little change in the optimal values: For CuH the only change was increase of the orbital exponent ζ on H-centered $1s$ orbital from 1.14 to 1.25, and for Cu the optimal value of the orbital exponent ζ on Cu-centered $4s$ orbital changed from 1.2 to 1.3. Also, H-centered $1s$ orbital with $c_{i,j} = 0.3$ and the orbital exponent $\zeta = 1.05$ was able to improve the Cu valence energy (recall that all H-centered orbitals on Cu turned out to have zero $c_{i,j}$'s when optimized using the original sampling).

Finally, we used these reoptimized values of the valence parameters in 8 runs, 500 blocks each[19], and obtained data in table 4.5. The valence energies are $E_{\mathrm{CuH}} = -0.873 \pm 0.009$ a.u. and $E_{\mathrm{Cu}} = -0.312 \pm 0.004$ a.u.

Taking into account the exact energy of the hydrogen atom, $E_{\mathrm{H}} = -0.5$ a.u., we

[18]The length and structure of the blocks as well the time steps were the same as described in the previous section.

[19]In this final run we decided to change the structure of the blocks: Instead of having 10 iterations of the core electrons followed by 100 iterations of the valence electrons we did 50 iterations of the core electrons followed by 50 iterations of the valence electrons in every block. Also, we decreased the time step for all core electrons 10 times.

E(CuH)	-0.843	-0.932	-0.917	-0.897	-0.914	-0.882
E(Cu)	-0.247	-0.256	-0.237	-0.249	-0.267	-0.238
Dipole moment	1.011	1.028	1.113	1.005	1.047	1.063

Table 4.4: CuH and Cu valence energies and CuH dipole moment (a.u.) obtained by the modified sampling for the valence parameters optimized using the original sampling (see Section 4.12 for more details). Results for six independent runs.

E(CuH)	-0.844	-0.910	-0.875	-0.874	-0.871	-0.900	-0.838	-0.869
E(Cu)	-0.321	-0.326	-0.305	-0.311	-0.300	-0.325	-0.311	-0.297
Dipole moment	1.783	1.782	1.785	1.798	1.793	1.793	1.808	1.810

Table 4.5: CuH and Cu valence energies and CuH dipole moment (a.u.) for the reoptimized valence parameters (see Section 4.12 for more details). Results for seven independent runs.

get the dissociation energy $D_e = E_{\mathrm{Cu}} + E_{\mathrm{H}} - E_{\mathrm{CuH}} = 0.61$ a.u. with standard error 0.008 a.u. This value differs from the experimental value 0.101 a.u. (see [13]) by approximately 40%.

Note, that the error bar of the dissociation energy is smaller than the sum of the error bars for the Cu and CuH valence energies. This is so because the error of the dissociation energy was estimated from 6 individual dissociation energies as obtained in the final runs rather than from the errors of the Cu and CuH energies. In this manner we take advantage of the fact that the sampling of the Cu and CuH valence enrgies is correlated (during the course of the simulation both systems have the same core electrons).

Value of the dipole moment, 1.794 ± 0.004 a.u., differs from the best CI calculation (see e.g. [64], $\mu = 1.2$) by about 50%.

Chapter 5

Comments, Suggestions, and Conclusions

In this chapter we would like to make a few suggestions about possible improvements of the presented $f \times g$ technique. Some of the suggestions we already tried. However, we did not explore any of them in details and it is therefore possible that in spite of our negative preliminary findings these may lead to interesting results. On the other hand, other suggestions we never really tried—they just came along our research and seem to be possible alternatives to what we were doing.

5.1 Configuration Interaction Approach

As already mentioned in Section 2.2, Hartree-Fock SCF with a wave function consisting of a single Slater determinant is not capable of adequately describing many electron atoms or molecules. A possible way of improving SCF is to take a linear combination of more Slater determinants—leading to CI.

A similar approach can be used in our $f \times g$ technique. However, we have to keep in mind that the essential part of the $f \times g$ approach is the core and valence energy separation. For the time being we do not see any other way of preserving the separation but keeping the same function f for all terms in the CI-like expansion and effectively modifying only the g function:

$$g = \sum_i \lambda_i \, g_i$$

where the functions g_i are products of two Slater determinants (for the spin up and spin down electrons) and a Jastrow factor (possibly, but not necessarily, the same for all g_i's).

As suggested by [65], this modification may be crucial in our case because the bond in CuH was found to have an important contribution from the mixture of Cu-centered $3d$ and H-centered $1s$ orbitals (see also [1]).

Being aware of the importance of the $3d$(Cu)-$1s$(H) interaction we implemented the described modification for a 2-term expansion of g. We used g_1 equal to the original g, i.e. containing the $4s$(Cu)-$1s$(H) interaction and g_2 with the $3d$(Cu)-$1s$(H) interaction. However, our optimization lead to $\lambda_2 = 0$, i.e. to the original $f \times g$ wave function.

We suspect the reason for this to be an 'incompatibility' of the functions f and g_2: The function f contains a $3d$(Cu) orbital for the core electrons and so does the function g_2—but for the valence ones! Also, the (normalized) product of f and g_2 does not contain a $4s$(Cu) orbital.

A possible solution is to decrease number of the core electrons from 28 to 26 and to increase number of the valence ones from 2 to 4 for CuH and from 1 to 3 for Cu. This gives us the possibility to deal with the $3d_0$(Cu) and $4s$(Cu) orbitals in the functions g_1 and g_2 and therefore removes the 'incompatibility'. However, we found that this solution brings another problem—the valence energy of both Cu and CuH decreased significantly[1] leading to very negative values for the dissociation energy. It is important to say that we did not fully optimized this form of the wave function. Therefore, it is possible that a further optimization would improve the dissociation energy.

5.2 Valence Energy

An interesting feature of VMC and DQMC algorithms is that they enable us in some sense—even though purely mathematical—to distinguish the electrons. This naturally

[1]Some preliminary results suggest the valence energy -2.5 a.u. for CuH and -1.6 a.u. for Cu.

leads to an attempt to define the valence electron(s) and corresponding valence energy—very interesting notions for chemists who are using the term 'valence electrons' extensively.

For long time we were trying to define the valence electrons[2] as the ones furthest from Cu nuclei. However, this definition may be difficult to extend to molecules with more heavy atoms. Therefore, we think that an alternative definition based on the (local) energy of individual electrons is worth exploring. According to this definition the valence electrons would be the ones with the least energy.

Also, the first steps have been done in exploring a possibility of branching[3] based on the valence energy.

5.3 Core Penetration

Occasionally, during the course of the simulation some valence electrons (y) become closer to the Cu nucleus than some core electrons (x). We call this effect a 'core penetration'.

Because of the full antisymmetry of the wave function $g(x, y)$ the valence electrons which penetrated the core have no reason to leave the core (increase their distance from the nuclei). However, their being close to the nuclei violates notion of the valence electrons as those further apart from the nuclei.

Therefore, we decided to modify the function g by a product of multiplicative factors

$$h(r_i) = \begin{cases} 0 & r_i < R - \epsilon \\ \left[\left(\frac{R-r_i}{\epsilon}\right)^2 - 1\right]^2 & R - \epsilon \leq r_i \leq R \\ 1 & R < r_i \end{cases}$$

for every valence electron x_i in the distance r_i from the nuclei, where R is the maximum

[2]Note that this definition does not apply to our $f \times g$ approach where the valence and core electrons are defined by the form of the $f \times g$ approximation.

[3]For DQMC, see Appendix A.

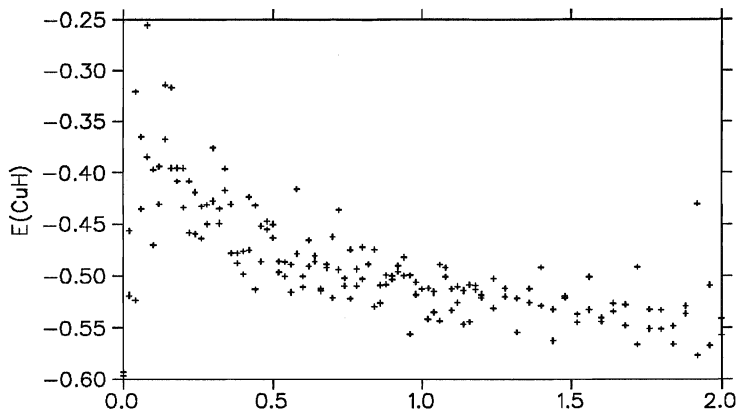

Figure 5.1: CuH valence energy (a.u.) vs. the 'smoothing' parameter ϵ for optimized parameters (every point represents average of 10 blocks; see Chapter 4.11 for more details)

distance of the core electrons y (having the same spin as x_i) from the nuclei. This essentially means that the valence electrons are subject only to g if they are further from the nuclei than the corresponding core electrons and they are not allowed to approach the core closer than $R-\epsilon$ (ϵ being a variational parameter). Interval distances between $R-\epsilon$ and R is used to damp the valence electrons conserving the continuity of the (modified) wave function g and its first derivatives.

However, as suggested in [66], this procedure may be a source of potential problems because for $\epsilon \rightarrow 0$ the expectation value for the kinetic energy is infinite.

According to Fig. 5.1 and 5.2 the optimal energy is achieved for larger values of the parameter ϵ. In our program we decided to use $\epsilon = 1.6$.

Note that the last two paragraphs are in slight contradiction: We would like to have the valence electrons x to be 'really' the valence ones, i.e. the outermost. That would imply small values of the 'smoothing' parameter ϵ. On the other hand we found that the (valence) energy has minimum for larger values of ϵ. For the time being we do not fully understand this.

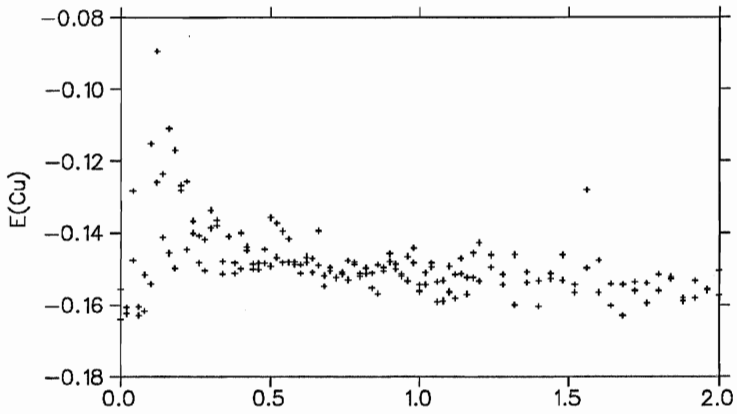

Figure 5.2: Cu valence energy (a.u.) vs. the 'smoothing' parameter ϵ for optimized parameters (every point represents average of 10 blocks; see Chapter 4.11 for more details)

5.4 Damped-Core Quantum MC

In [37] the so called damped-core quantum MC method is described for treatment for large-Z (many electron) systems. This method provides ionization potentials and electron affinities of C, Si and Ge which are in agreement with experiment.

The method can be described as a simplified version of our $f \times g$ approach for $g = g(y)$ only, i.e. the valence electrons do not depend on the core ones explicitly via the wave function. This assumption leads to some technical simplifications (e.g. the variance term we had problems to deal with vanishes). At the same time the method uses idea of branching based on the valence energy (see the previous section).

It would be of interest to apply this approach to see if it can be successful for a molecule like CuH.

5.5 Conclusions

In this work we presented a new $f \times g$ approach to calculation of the dissociation energy and applied it to CuH. Even though in its present form the approach is not competitive with the best theoretical calculations recovering more than 90% of the dissociation energy, the error of our results is comparable with the error of SCF calculations or simple pseudpotential methods (see table 5.1).

Also, our estimate of the dipole moment differs from the most up-to-date CI calculation (see [64]) by approximately 50%. However, as table 5.1 suggests, this is comparable with the results provided by both CAS SCF and pseudopotential methods.

We think that the qualitative agreement between our results and the ones obtained by the pseudopotential methods is due to some similarities of the two techniques, namely that both approaches rely on the valence electrons and their energy rather than on the total energy. The difference is that in a typical pseudopotential calculation one has an explicit form of the potential available, whereas in our case it is not so. Instead, we dynamically simulate the core electrons and the valence ones are adjusted accordingly.

The error of our results we attribute—in addition to the error of the $f \times g$ approximation itself—mainly to the small basis set used. Also, we suspect the small basis set coupled with basis set superposition error (which we treated by the counterpoise method) to be the main cause of the dissociation energy being too large.

At present time we see two possible orientations of further research: to explore the problem with the variance term in our approach by applying the technique to a smaller system[4], and/or to explore the CI approach further.

[4]It is natural to expect that for a smaller system it is easier to distinguish between technical and principal problems.

Table 5.1: Nonrelativistic estimates of dissociation energy and dipole moment for CuH

Method	D_e (a.u.)	μ (a.u.)
$f \times g$	0.610(6)	1.794(4)
Previous work		
SCF[a]	0.051	
Pseudopotential[b]	0.078	
Pseudopotential[c]	0.096	1.715
CAS SCF[d]	0.071	1.705
MP2[d]	0.095	1.209
best CI[e]		1.2
exp.[f]	0.101	

[a] From ref. [67]
[b] From ref. [68]
[c] From ref. [69]
[d] From ref. [1]
[e] From ref. [64]
[f] Recommended value from [13] according to [14].

Appendix A

Diffusion Quantum Monte Carlo

In this chapter we describe the diffusion quantum MC method, which picks up from our discussion of VMC (see page 15). Also, idea of 'branching' in DQMC could be used in our $f \times g$ approach (probably 'branching' based on the valence energy) and could possibly improve the results.

First we describe a simple case, when the wave function of the ground state we are interested in is positive everywhere, e.g. for a system of identical bosons.

Then we proceed to systems of identical fermions, which are more complicated because in this case the ground state wave function is positive in some regions and negative in others[1].

There are many different techniques addressing the sign problem. We will describe only a fixed-node approximation, which is relevant for the previous discussion about VMC.

A.1 Basic Idea

We start from the time dependent Schrödinger equation in atomic units

$$i\,\frac{\partial \Psi(x,t)}{\partial t} = \left(\hat{H} - E_T\right)\Psi(x,t) \tag{A.1}$$

[1] An exception is a system of two electrons which have opposite spins and the space part of the wave function is therefore positive everywhere.

where x denotes coordinates of all electrons,

$$\hat{H} = -\frac{1}{2}\triangle + V(x)$$

is the Hamiltonian, $V(x)$ is the potential energy operator and E_T is an arbitrary constant[2].

Substitution $it = \tau$ in equation (A.1) gives

$$\frac{\partial \Psi(x,\tau)}{\partial \tau} = -\left(\hat{H} - E_T\right)\Psi(x,\tau) \tag{A.2}$$

or, in integral form,

$$\Psi(x,\tau) = \int G(x,x',\tau)\,\Psi(x',0)\,dx' \tag{A.3}$$

where $G(x',x,\tau)$ is the Green function defined by

$$\frac{\partial G(x,x',\tau)}{\partial \tau} = -\left(\hat{H} - E_T\right)G(x,x',\tau) \tag{A.4}$$

and the initial condition

$$G(x,x',0) = \delta(x-x') \tag{A.5}$$

In the following we need a formal expressions for the Green function, namely

$$G(x,x',\tau) = \sum_n e^{-(E_n - E_T)\tau}\psi_n^*(x')\,\psi_n(x) \tag{A.6}$$

where ψ_n are normalized eigenfunctions of $\hat{H}$ and E_n are corresponding eigenvalues

$$\hat{H}\psi_n = E_n\psi_n$$

and we assume that ψ_0 and E_0 correspond to the ground state.

Using (A.6) and (A.3) we get

$$\Psi(x,\tau) = \sum_n a_n e^{-(E_n - E_T)\tau}\psi_n(x) \tag{A.7}$$

[2]Note that an arbitrary constant may be subtracted from the Hamiltonian without having any influence on the solution of the Schrödinger equation. Reason for introducing this constant will be obvious later.

where

$$a_n = \int \psi_n^*(x)\, \Psi(x,0)\, dx$$

Since E_0 is the ground state energy, the first term in (A.7) will dominate[3] for large τ, i.e. $\Psi(x,\tau) \sim \psi_0(x)$ for $\tau \to \infty$. By a proper choice of E_T ($E_T = E_0$) we can achieve ψ_0 be a stationary solution.

Let's assume that the Green function $G(x,x',\tau)$ is known for some time step $\tau = \tau_0$. Then, the repetitive usage of (A.3) enables us to compute $\Psi(x,\tau \to \infty)$, i.e. the ground state ψ_0.

If both the Green function $G(x,x',\tau_0)$ and ground state ψ_0 are positive then we can represent $\Psi(x,\tau)$ by a sample of points (walkers) $\{x_1, x_2, \ldots\}$ drawn from the probability distribution function $\Psi(x,\tau)$.

A short time approximation of the Green function is derived in [43]. The result is:

$$G(x,x',\tau_0) \approx \frac{1}{(2\pi\tau_0)^{3N/2}} \exp\left[-\frac{(x-x')^2}{2\tau_0}\right] e^{-(V(x')-E_T)\tau_0} \tag{A.8}$$

where N is the number of particles involved (x is a $3N$ dimensional vector).

Hence, the equation (A.3) can be interpreted using the ensemble of walkers:

- The first part of the Green function (A.8) represents diffusion from x' to x. We can simulate this part by random move from x' to x with the probability density function

$$\frac{1}{(2\pi\tau_0)^{3N/2}} \exp\left[-\frac{(x-x')^2}{2\tau_0}\right]$$

- The second part of the Green function represents growth (or death) of walkers, and it can be simulated by branching, i.e. creating

$$\mathrm{INT}\left\{e^{-(V(x')-E_T)\tau_0} + \mathrm{RAND}\right\}$$

[3] Assuming $a_0 \neq 0$.

copies of the walker[4] at x', where RAND is a random number between 0 and 1 and INT(x) is the largest integer less or equal to x.

Due to the branching it is possible that the total number of walkers tends to increase or decrease systematically. In that case we have to adjust value of E_T so that a stable situation is achieved[5]. Then, the value E_T is an estimate of the ground state energy.

A.2 Fermion Case

Due to the antisymmetry of a fermion wave function, the interpretation of $\Psi(x,\tau)$ as a probability density distribution is no longer possible (Ψ is positive at some regions and negative at others).

However, it is possible to rewrite (A.2) into a new form:

$$\frac{\partial f(x,\tau)}{\partial \tau} = \frac{1}{2} \triangle f(x,\tau) - \nabla \left[f(x,\tau) F(x) \right] - \left(E_l(x) - E_T \right) f(x,\tau) \tag{A.9}$$

where the new function

$$f(x,\tau) = \Psi(x,\tau)\, \psi_T(x) \tag{A.10}$$

the velocity

$$F(x) = \frac{\nabla \psi_T(x)}{\psi_T(x)}$$

and the local energy

$$E_l = \frac{\hat{H} \psi_T(x)}{\psi_T(x)}$$

are defined using a trial function $\psi_T(x)$.

If we were able to find a trial function $\psi_T(x)$, nodes of which are exactly the nodes of the exact solution of the Shrödinger equation, the new function $f(x,\tau)$ would be positive

[4] Or deleting the walker if the above expression turns out to be zero.

[5] Of course, we can not remove the *statistical* fluctuations of the number of walkers.

everywhere and we could solve (A.9) in a manner very similar to that from the previous section.

The only nontrivial modification would be due to the term

$$\nabla [f(x,\tau)\, F(x)]$$

This term modifies the approximation (A.8) in the following way (see [43, 39, 26, 60]):

$$G(x,x',\tau_0) \approx \frac{1}{(2\pi\tau_0)^{3N/2}} \exp\left[-\frac{(x-x'-\tau_0\, F(x'))^2}{2\tau_0}\right] e^{-(E_l(x')-E_T)\tau_0} \qquad \text{(A.11)}$$

and therefore can be simulated by moving the walker from x' to $x' + \tau_0\, F(x')$ (followed by a diffusion step and branching).

However, the exact location of nodes of the wave function ψ_0 is not known.

Therefore, we have to choose $\psi_T(x)$ be an approximate wave function which doesn't describe the nodes properly. And because in our simulation $f(x,\tau)$ is not negative[6], the trial function effectively imposes nodes of $\psi_T(x)$ on the final solution ψ_0 (see (A.10)).

In this way DQMC for fermions has not only a τ bias due to the approximation involved in (A.11), but also a bias due to the incorrect nodes. It has been shown (see [60, 70]) that the fixed node bias is positive, i.e. one obtains an upper bound to the ground state energy. The τ bias is usually dealt with by using more values of τ and then extrapolating to zero τ.

[6]It is treated as a probability distribution function.

Appendix B

Energy and Dipole Moment in $f \times g$ Approach

In this Appendix we derive the expressions (4.5) and (4.6) for the total energy E

$$E = \frac{\int \psi^* \hat{H} \psi \, dV}{\int \psi^* \psi \, dV}$$

with the wave function $\psi(x, y)$

$$\psi(x, y) = f(x) \frac{g(x, y)}{\sqrt{\int g^2(x, y) \, dy}}$$

and the Hamiltonian

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^{N_e} \triangle_i + \hat{V}$$

where N_e is the total number of electrons involved and $\hat{V}$ is the potential energy operator.

Also, we derive the formula used to evaluate the dipole moment of CuH.

B.1 Potential Energy

Dealing with the potential energy is rather simple: First we write V as a sum of three terms

$$\hat{V} = V_c(x) + V_v(x, y) + V_{nn}$$

where $V_c(x)$ depends on the core electrons x only, i.e. this term includes interactions between the core electrons, and between the core electrons and nuclei. Other interactions involving some electron(s), i.e. the interactions of valence electrons with themselves, with the core electrons and with the nuclei, are included in $V_v(x, y)$. The term V_{nn} represents

repulsion between nuclei for CuH (or is not present for Cu) and is independent on both x and y.

Using this separation, the potential energy is

$$U = \frac{\int \psi^* \hat{V} \psi \, dV}{\int \psi^* \psi \, dV} = \frac{\int f^2(x) \frac{g^2(x,y)}{\int g^2(x,y)\, dy} (V_c(x) + V_v(x,y) + V_{nn}) \, dx \, dy}{\int f^2(x) \frac{g^2(x,y)}{\int g^2(x,y)\, dy} \, dx \, dy}$$

and after integrating over x and y where possible

$$U = \frac{\int f^2(x)\, V_c(x)\, dx}{\int f^2(x)\, dx} + \frac{\int \frac{\int g^2(x,y)\, V_v(x,y)\, dy}{\int g^2(x,y)\, dy} f^2(x)\, dx}{\int f^2(x)\, dx} + V_{nn}$$

In the last expression it is easy to recognize averaging over the core and valence configurations

$$U = \langle V_c \rangle_c + \langle \langle V_v \rangle_v \rangle_c + V_{nn} \tag{B.1}$$

B.2 Kinetic Energy

Writing the kinetic energy operator

$$\hat{T} = -\frac{1}{2} \sum_{i=1}^{N_e} \triangle_i$$

in form

$$\hat{T} = -\frac{\triangle_x}{2} - \frac{\triangle_y}{2}$$

gives

$$\begin{aligned} T &= \frac{\int \psi^* \hat{T} \psi \, dV}{\int \psi^* \psi \, dV} \\ &= \frac{\int f(x) \frac{g(x,y)}{\sqrt{\int g^2(x,y)\, dy}} (-\frac{\triangle_x}{2} - \frac{\triangle_y}{2}) f(x) \frac{g(x,y)}{\sqrt{\int g^2(x,y)\, dy}} \, dx \, dy}{\int f^2(x) \frac{g^2(x,y)}{\int g^2(x,y)\, dy} \, dx \, dy} \end{aligned} \tag{B.2}$$

The valence contribution to the kinetic energy, i.e. the term with $-\triangle_y/2$, simplifies to

$$T_v = \frac{\int f^2(x) \frac{g^2(x,y)}{\int g^2(x,y)\,dy} \left(-\frac{1}{2}\frac{\triangle_y g(x,y)}{g(x,y)}\right) dx\,dy}{\int f^2(x)\,dx}$$

or

$$T_v = \left\langle -\frac{1}{2}\frac{\triangle_x g(x,y)}{g(x,y)} \right\rangle_{v,c} \tag{B.3}$$

To get the core contribution to the kinetic energy, i.e. term with $-\triangle_x/2$ in (B.2), we first express $\triangle_x \psi(x,y)$ in terms of $f(x)$ and $g(x,y)$:

$$\begin{aligned}
&\triangle_x \left[f(x) \frac{g(x,y)}{\sqrt{\int g^2(x,y)\,dy}} \right] \\
&\quad = \nabla_x \left[\nabla_x f \frac{g}{\sqrt{\int g^2 dy}} + f \frac{\nabla_x g}{\sqrt{\int g^2 dy}} - f \frac{g}{(\int g^2 dy)^{3/2}} \int g\, \nabla_x g\, dy \right] \\
&\quad = \triangle_x f \frac{g}{\sqrt{\int g^2 dy}} + 2\,\nabla_x f \frac{\nabla_x g}{\sqrt{\int g^2 dy}} - 2\,\nabla_x f \frac{g}{(\int g^2 dy)^{3/2}} \int g\, \nabla_x g\, dy \\
&\qquad + f \frac{\triangle_x g}{\sqrt{\int g^2 dy}} - 2f \frac{\nabla_x g}{(\int g^2 dy)^{3/2}} \int g\, \nabla_x g\, dy + (2+1) f \frac{g}{(\int g^2\, dy)^{5/2}} \left(\int g\, \nabla_x g\, dy \right)^2 \\
&\qquad - f \frac{g}{(\int g^2 dy)^{3/2}} \int (\nabla_x g)^2\, dy - f \frac{g}{(\int g^2 dy)^{3/2}} \int g\, \triangle_x g\, dy
\end{aligned}$$

Now we should multiply the above expression by ψ and integrate over x and y. However, it is easy to see that after integration over y the second term will cancel the third one, the fourth term will cancel the eight one, and the fifth term will cancel factor 2 in the sixth term. Therefore, the only terms with a nonzero contribution to the kinetic energy are

$$\begin{aligned}
&\triangle_x f \frac{g}{\sqrt{\int g^2 dy}} + f \frac{g}{(\int g^2\, dy)^{5/2}} \left(\int g\, \nabla_x g\, dy \right)^2 - f \frac{g}{(\int g^2 dy)^{3/2}} \int (\nabla_x g)^2\, dy \\
&\qquad = \frac{\triangle_x f}{f} \frac{f\,g}{\sqrt{\int g^2\, dy}} + \left\langle \frac{\nabla_x g}{g} \right\rangle_v^2 \frac{f\,g}{\sqrt{\int g^2\, dy}} - \left\langle \left(\frac{\nabla_x g}{g} \right)^2 \right\rangle_v \frac{f\,g}{\sqrt{\int g^2\, dy}}
\end{aligned}$$

and the core contribution to the kinetic energy, simplifies to

$$T_c = \left\langle -\frac{1}{2}\frac{\triangle_x f(x)}{f(x)} \right\rangle_c + \frac{1}{2}\langle \mathrm{Var} \rangle_c \tag{B.4}$$

where

$$\mathrm{Var} = \left\langle \left(\frac{\nabla_x g(x,y)}{g(x,y)} \right)^2 \right\rangle_v - \left\langle \frac{\nabla_x g(x,y)}{g(x,y)} \right\rangle_v^2 \tag{B.5}$$

Combination of the valence contribution (B.3) and the core contribution (B.4) to the kinetic energy gives

$$T = \left\langle -\frac{1}{2}\frac{\triangle_x f(x)}{f(x)} \right\rangle_c + \left\langle -\frac{1}{2}\frac{\triangle_x g(x,y)}{g(x,y)} \right\rangle_{v,c} + \frac{1}{2}\langle \mathrm{Var} \rangle_c \tag{B.6}$$

B.3 The Total Energy

From the above derived expressions for the potential energy (B.1) and for the kinetic energy (B.6) it is obvious that the total energy is given by (4.5) with Var defined by (4.6).

B.4 Dipole Moment

To deal with the dipole moment we first fix the coordinate system: Cu nucleus is chosen to be at the origin of the coordinate system and H nucleus on the negative z axis. With this choice the only nonzero component of the dipole moment is the z component (the other two components are zero because of the rotational symmetry of CuH molecule along z axis).

Therefore, the dipole moment is given by

$$\mu = \frac{\int \psi^* \hat{d} \psi \, dV}{\int \psi^* \psi \, dV}$$

where the operator corresponding to the dipole moment is equal to the sum of z coordinates of all electrons and can be written as a sum of two sums—over the core and the

valence electrons:

$$\hat{d} = \sum_{\text{all}} z_i = \sum_{\text{core}} z_i + \sum_{\text{val.}} z_i$$

A calculation similar to the one leading to the equation (B.1) gives

$$\mu = \left\langle \sum_{\text{core}} z_i \right\rangle_c + \left\langle \left\langle \sum_{\text{val.}} z_i \right\rangle_v \right\rangle_c$$

However, because of the symmetry of the core (function f) the first term is equal to zero and the dipole moment is given by

$$\mu = \left\langle \left\langle \sum_{\text{val.}} z_i \right\rangle_v \right\rangle_c$$

Appendix C

Cusp Conditions for the Jastrow Factor

In this Appendix we derive the cusp conditions for the Jastrow parameters $a_{i,j}$ as described in Section 4.8.

For the wave function ψ of the form

$$\psi = f \exp\left(\sum_{i<j} J_{i,j}(r_{i,j})\right) \tag{C.1}$$

the kinetic energy T is

$$\begin{aligned} T &= -\frac{1}{2}\frac{\triangle\psi}{\psi} \\ &= -\frac{1}{2}\sum_k \frac{\triangle_k f}{f} - \sum_{i\neq k} \frac{\nabla_k f}{f}\frac{\partial J_{i,k}}{\partial r_{i,k}}\frac{\vec{r}_{i,k}}{r_{i,k}} - \frac{1}{2}\sum_k \left(\sum_{i\neq k}\frac{\partial J_{i,k}}{\partial r_{i,k}}\frac{\vec{r}_{i,k}}{r_{i,k}}\right)^2 \\ &\quad -\frac{1}{2}\sum_{i\neq k}\frac{\partial^2 J_{i,k}}{\partial r_{i,k}^2} - \sum_{i\neq k}\frac{\partial J_{i,k}}{\partial r_{i,k}}\frac{1}{r_{i,k}} \end{aligned} \tag{C.2}$$

where $\vec{r}_{i,k} = \vec{r}_k - \vec{r}_i$.

Now let us consider two electrons, i and k, approaching each other.

If the two electrons are of the opposite spin, f is not necessarily zero and it is only the last term in (C.2) which diverges as $1/r_{i,k}$. However, in the electron-electron potential energy

$$V_{ee} = \frac{1}{2}\sum_{i\neq k}\frac{1}{r_{i,k}}$$

we have a term with the same behavior. Therefore, if we choose

$$\left.\frac{\partial J_{i,k}}{\partial r_{i,k}}\right|_{r_{i,k}=0} = \frac{1}{2} \qquad \text{for } i,\ k \text{ of the opposite spin} \tag{C.3}$$

the two divergent terms will cancel each other and the total energy will be well-behaved.

If the two electrons are of the same spin, f vanishes as $r_{i,k} \to 0$ (f is antisymmetric with respect to any two spin-like electrons). For small $r_{i,k}$ we can write

$$f \approx \vec{c}\vec{r}_{i,k}$$

where $\vec{c}$ is a constant vector.

This approximation gives no singularities in $\frac{\triangle_k f}{f}$ and

$$\frac{\nabla_k f}{f} \approx \frac{\vec{c}}{\vec{c}\vec{r}_{i,k}}$$

Hence, by choosing

$$\left.\frac{\partial J_{i,k}}{\partial r_{i,k}}\right|_{r_{i,k}=0} = \frac{1}{4} \qquad \text{for } i,\ k \text{ of the same spin} \tag{C.4}$$

we can make sure that the $1/r_{i,k}$ singularity of the second and the last term in (C.2) will cancel with the corresponding term from the potential energy and the total energy will be well-behaved.

Equations (C.3) and (C.4) give the general form of the cusp conditions for the Jastrow factor $J_{i,k}$.

The Jastrow factor in our simulations is

$$J_{i,k} = \frac{a_{i,k}\, r_{i,k}}{1 + b_{i,k} r_{i,k}}$$

with the derivative

$$\left.\frac{\partial J_{i,k}}{\partial r_{i,k}}\right|_{r_{i,k}=0} = a_{i,k}$$

Therefore, the conditions (C.3) and (C.4) lead to

$$a_{i,k} = \begin{cases} 1/4 & \text{if } i \text{ and } k \text{ are of like spin} \\ 1/2 & \text{if } i \text{ and } k \text{ are of unlike spin} \end{cases}$$

as discussed in Section 4.8.

Appendix D

Stale Configurations

As mentioned in Section 3.6, configurations which are near the nodal surface of the wave function ψ (for which $\psi(x) \approx 0$) tend to be the stale configurations, i.e. the acceptance probability (3.11) of a move from x to x' approaches 0 for $x \to x_0$, where $\psi(x_0) = 0$.

In this appendix we present a proof of this statement according to [71, Appendix III].

First we recall that the acceptance probability (3.11) of a move from x to x' is given by

$$\min\left(1, \frac{\psi^2(x')}{\psi^2(x)} \frac{G(x, x', \tau)}{G(x', x, \tau)}\right) \tag{D.1}$$

where

$$G(x', x, \tau) = \frac{1}{(2\pi\tau)^{3N/2}} \exp\left[-\frac{(x' - x - \tau F(x))^2}{2\tau}\right]$$

$$x' \approx x + \tau F(x)$$

(ignoring the diffusion term independent on both x and x') and

$$F(x) = \frac{\nabla\psi(x)}{\psi(x)}$$

Therefore, the acceptance probability for any fixed τ is given by

$$\min\left(1, \frac{\psi^2(x')}{\psi^2(x)} \exp\left[-\frac{\tau}{2}\left(F(x) + F(x')\right)^2\right]\right)$$

For $\psi(x) \to 0$ the last expression is dominated by $F(x)$ and the acceptance probability is approximately

$$\min\left(1, \frac{\psi^2(x')}{\psi^2(x)} \exp\left[-\frac{A^2}{\psi^2(x)}\right]\right)$$

where A is a real constant. Finally, using

$$\lim_{y \to 0^+} \frac{e^{-A^2/y}}{y} = 0$$

and the fact that the wave function ψ is bounded we conclude that the acceptance probability approaches 0 for $\psi(x) \to 0$.

Appendix E

Matrix Algebra

In the first part of this Appendix we present formulas for calculation of the local energy, wave function and the drift as described in [56]. The second part describes a trick used to improve efficiency of the calculation for $f \times g$ approach. Finally, in the last section of this Appendix we explain why the tricks described do not have the expected effect.

E.1 The Local Energy

We assume a trial function of the form $\Psi = \psi^{\uparrow}\psi^{\downarrow}\psi^{c}$, where $\psi^{\uparrow}$ is a Slater determinant of molecular orbitals (linear combination of atomic orbitals in our case) for the spin up electrons, $\psi^{\downarrow}$ the analog for the spin down electrons, and ψ^{c} is the Jastrow correlation function.

The local energy $E_l = \hat{H}\Psi/\Psi$ is computed as follows:

$$E_l = -\frac{1}{2}\left(g^{\uparrow} + g^{\downarrow} + 2\sum_i \left(F_i^{\uparrow}\cdot F_i^{c\uparrow} + F_i^{\downarrow}\cdot F_i^{c\downarrow}\right) + g^{c}\right) + V \qquad \text{(E.1)}$$

where $g^{\uparrow} = \triangle_N\psi^{\uparrow}/\psi^{\uparrow}$, $F_i^{\uparrow} = \nabla_i\psi^{\uparrow}/\psi^{\uparrow}$, and V is the potential energy. Also, ∇_i is the 3-dimensional gradient operator corresponding to the i^{th} electron, and $\triangle_N$ is the $3N$ dimensional Laplace operator, where N is the number of spin up ($\uparrow$) or down ($\downarrow$) electrons. (In general $N^{\uparrow} \neq N^{\downarrow}$, but we omit the superscript on N.)

The following formulas are equally valid for both superscripts $\uparrow$ and $\downarrow$. Therefore, we omit the superscripts from our notation.

Let A be the $J \times N$ matrix of values of J atomic orbitals, each evaluated N times,

corresponding to the locations of the individual electrons. Let B and D be the matrix of the corresponding gradients and $\triangle_i$ values of the atomic orbitals, respectively. (B has three components for each entry; i.e. it is a $J \times N \times 3$ matrix.) Finally, let C be the $N \times J$ matrix of molecular orbital coefficients. Thus CA is the Slater matrix of N molecular orbitals evaluated at each electron's position, and $\psi = \det(CA)$.

One can easily compute the quantities appearing in (E.1) by calculating the inverse of the Slater matrix $(CA)^{-1}$, and some simple matrix algebra [72]. In particular,

$$g = \sum_{i,j} (CD)_{i,j} (CA)^{-1}_{j,i} \tag{E.2}$$

and

$$F_i = \sum_{j} (CB)_{j,i} (CA)^{-1}_{i,j} \tag{E.3}$$

Here we again stress that B has the extra dimension of 3, corresponding to the x, y, z components of the gradient.

E.2 Simplification

Let a matrix M consists of 4 matrices A, B, C, and D:

$$M = \begin{pmatrix} A & C \\ D & B \end{pmatrix}$$

It is easy to verify, that the inverse of M is

$$M^{-1} = \begin{pmatrix} X & Z \\ T & Y \end{pmatrix}$$

where

$$Y = \left(B - DA^{-1}C\right)^{-1}$$

$$Z = -A^{-1}CY$$

$$T = -YDA^{-1}$$

$$X = A^{-1} - A^{-1}CT$$

As suggested by [57], the above formulas can be very useful for evaluation of the Slater matrix inverse, $(CA)^{-1}$ in (E.2) and (E.3), in our $f \times g$ approach. The key idea is that the Slater matrix for g contains the Slater matrix for f in the same way as M contains A. Therefore, we can calculate the Slater matrix inverse for f using the above matrix algebra and the Slater matrix inverse for g.

The advantage of this becomes clear when we realize, that the number of steps (multiplications and additions) a computer has to do to obtain a matrix inverse is approximately proportional to the matrix's dimension (in our case the number of the electrons) cubed. Therefore, it is more efficient to use the above described algebra, which at the most involves multiplication of 'lean' matrices[1] and is therefore computationally faster, rather than performing the full matrix inversion. Gain of this approach is even more evident if we recall that in our algorithm every step of the core electrons is followed by many steps of the valence electrons—and it is only the core move which requires recalculation of the Slater matrix inverse for f.

Another trick used in our program to evaluate the Slater determinant is based on the following identity:

$$\begin{aligned}
\det \begin{pmatrix} A & C \\ D & B \end{pmatrix} &= \det \left[\begin{pmatrix} A & C \\ D & B \end{pmatrix} \begin{pmatrix} 1 & -A^{-1}C \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -A^{-1}C \\ 0 & 1 \end{pmatrix}^{-1} \right] \\
&= \det \left[\begin{pmatrix} A & 0 \\ D & B - DA^{-1}C \end{pmatrix} \begin{pmatrix} 1 & -A^{-1}C \\ 0 & 1 \end{pmatrix}^{-1} \right] \\
&= \det A \, \det \left(B - DA^{-1}C \right)
\end{aligned}$$

[1]Matrices with only one (or none) dimension equal to the number of the core electrons. The other dimension(s) is (are) given by the number of the valence electrons, which is significantly smaller, e.g. 14 versus 2 for the spin up electrons in CuH.

As in the previous case, the Slater determinant for the core electrons $\det A$ is evaluated after every core move. However, after each valence move we need to evaluate only $\det(B - DA^{-1}C)$, what is computationally much less demanding than evaluation of the full determinant.

E.3 Catch

As already mentioned in Chapter 4, one of the two main reasons leading to the approximation (4.1) was the possibility to make the calculation efficient using the tricks described in the previous section of this Appendix. It was expected that the tricks would enable to move the valence electrons significantly faster than the core ones.

In this section we show that this is not the case and that a move of the valence electrons takes approximately the same CPU time as a move of the core electrons (as explained below, the valence electrons can move in slightly less or slightly more CPU time than the core ones).

In the following analysis we consider evaluation of only one Slater determinant involving n_c core electrons[2] and n_v valence electrons[3], and its first (the drift F) and second (the kinetic energy) derivatives.

The two most time consuming calculations for the core electrons are evaluation of the Slater determinant (f), which needs about n_c^3 multiplications[4], and evaluation of its first derivatives, which takes approximately $3\,n_c^2$ multiplications[5].

For the valence electrons the two most time consuming calculations (involving g) are evaluation of the drift, which takes about $3\,n_v\,n_c^2$ multiplications[6], and evaluation of the

[2]For both functions f and g.

[3]For the function g only.

[4]That is the number of multiplications needed to compute determinant of a square matrix $n_c \times n_c$.

[5]See (E.3), where both indices i and j have values between 1 and n_c and every vector F_i has three components corresponding to x, y and z components of the drift.

[6]To get $n_c \times n_c$ entries in in the submatrix of CB in (E.3), each one of which is obtained from 1 core

kinetic energy, which needs approximately $n_v\, n_c^2$ multiplications[7].

Therefore, the ratio of the valence and core time per move is

$$\frac{T_{valence}}{T_{core}} = \frac{3\, n_v\, n_c^2 + n_v\, n_c^2}{n_c^3 + 3\, n_c^2} = \frac{4\, n_v}{n_c + 3}$$

where in our program $n_c = 14$ and the smallest value for n_v is 3.5[8] giving the best ratio $T_{valence}/T_{core} \approx 0.82$. For the maximum number of the valence orbitals we used (9) $n_v = 18$ and the valence move is actually slower than the core move.

Of course, this does not mean that the tricks used are of no value in our calculation. If we did not use them the valence move calculation would be much slower than it is now and the ratio $T_{valence}/T_{core}$ would become much worse. It is so because then the valence move calculation would be computationally the same as for the core move, except that the number of electrons would increase from n_c to $n_c + n_v$ and gave the ratio

$$\frac{T_{valence}}{T_{core}} = \frac{(n_c + n_v)^3 + 3\,(n_c + n_v)^2}{n_c^3 + 3\, n_c^2}$$

For $n_c = 14$ and $n_v = 3.5$ the ratio would be 1.88, i.e. worse than the previous case with the ratio 0.82

orbital (neglected) and n_v valence orbitals.

[7]Analogical to the drift but evaluating the entries in CD in (E.2).

[8]There are 3 valence orbitals on CuH (Cu-centered $4s$ and H-centered $1s$ and $2p_0$) and on the average 0.5 valence orbitals on Cu (Cu-centered $4s$ orbital for the spin up electrons and nothing for the spin down electrons).

Appendix F

Improved Sampling

As already mentioned in Section 3.6, the Metropolis algorithm described in Chapter 3 does not handle the nodal regions of the wave function properly. In this Appendix we describe a possible solution of this problem suggested in [42].

First of all, to visualize the problem of the original algorithm we took a simple one dimensional wave function

$$\psi = c\,\frac{x}{x^2+1}\,\mathrm{e}^{-x^2}$$

where c is a normalization constant, and simulated the probability distribution ψ^2. Results are depicted on Fig. F.1: Solid line represents the distribution ψ^2, and the bins are obtained from a run with 1,000,000 steps and time step $\tau = 1.0$. From this figure it is clear that the original algorithm is not sampling the nodal region properly. Even though for smaller τ's the sampling of the nodal region improves (size of the uncovered region decreases), the deficiency of the algorithm is obvious.

In [42] it is suggested to replace the drift $F = \frac{\nabla_i \psi}{\psi}$ of every individual electron in equation (3.10) and in the related equations with

$$F' = \frac{-1+\sqrt{1+2a\tau F^2}}{a\tau F^2} F \tag{F.1}$$

With this choice, $F' \approx F$ for small $a\tau F^2$, i.e. the short drift moves are almost unchanged. However, for very large F—indicating proximity of the nodal region—the drift moves have their length limited to $\sqrt{2\tau a}$.

Motivation for this modification is as follows: In a close proximity of the nodal surface

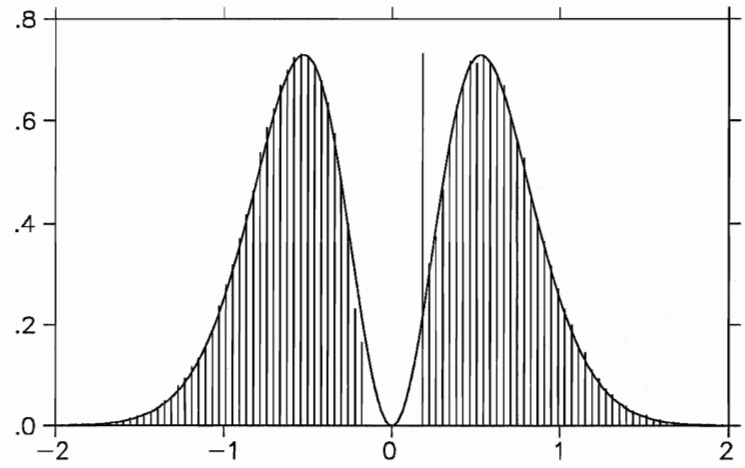

Figure F.1: Original sampling

the drift is given by

$$F(x) = \frac{\nabla\psi(x)}{\psi(x)} \approx \frac{\hat{R}_\perp}{R_\perp}$$

where $R_\perp$ is the distance to the node and $\hat{R}_\perp$ is a unit vector orthogonal to the nodal surface and pointing from the nodal surface to x. Therefore,

$$dR_\perp = \frac{1}{R_\perp}\, d\tau$$

or, after integration over the time interval τ,

$$R_\perp(\tau) - R_\perp(0) = \sqrt{R_\perp^2(0) + 2\tau} - R_\perp(0) = F'\tau$$

by definition of F'. Now, using the approxiamtion $F \approx \hat{R}_\perp / R_\perp$ and going back to vectors, we get the equation (F.1) with $a = 1$.

In order to compensate for largeness of F near a nucleus, authors of the modification propose to use a position dependent parameter a:

$$a = \frac{1}{2}(1 + \hat{F}\,\hat{z}) + \frac{Z^2 z^2}{10(4 + Z^2 z^2)}$$

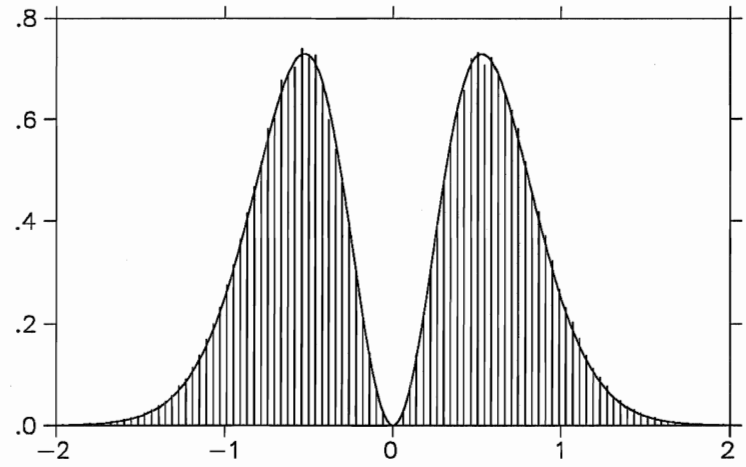

Figure F.2: New sampling

where $\hat{z}$ is a unit vector from the nearest nucleus with charge Z to the electron, which is at distance z, while $\hat{F}$ is a unit vector in the direction of the single electron velocity $\frac{\nabla_i \psi}{\psi}$.

Effect of the modification on our one dimensional example is remarkable: As illustrated on Fig. F.2, which was obtained under the same conditions as Fig. F.1 except for the modified drift with $a = 1$, the improved algorithm seems to have no problems with handling the nodal region of the wave function ψ.

Appendix G

A Simple SCF Model of CuH

In order to find out whether it is possible to obtain binding between Cu and H using only Cu centered $4s$ and H centered $1s$ orbitals and how strong such a bond is we decided to make a simple SCF calculation using only the two orbitals.

For this purpose we approximated the CuH molecule by two point charges representing Cu^+ and H^+ ions. The distance between the ions was chosen to be[1] $R = 2.764$ a.u., i.e. the experimental bond distance for CuH (see [62]). Then we varied the orbital exponents ζ of the two orbitals, calculated the coefficients a and b in the linear combination of the two orbitals, $a\,4s(\mathrm{Cu}) + b\,1s(\mathrm{H})$, using SCF and finally evaluated the energy of the two electrons with opposite spins occupying the molecular orbital.

The obtained energy surface was projected on the energy vs. $4s(\mathrm{Cu})$ ζ (see fig. G.1) and energy vs. $1s(\mathrm{H})$ ζ (see fig. G.2) planes. These figures indicate that there is a minimum on the energy surface for $4s(\mathrm{Cu})$ $\zeta = 0.87$ and $1s(\mathrm{H})$ $\zeta = 0.50$. The optimal energy turns out to be $E_{\mathrm{CuH}} = -0.9852$.

A similar calculation was done for Cu. However, in this case there was only one electron occupying the orbital $a\,4s(\mathrm{Cu}) + b\,1s(\mathrm{H})$ and the $1s(\mathrm{H})$ orbital was treated as a ghost orbital, i.e. without any charge in its center.

The energy surface (see fig. G.3 and G.4) predicts minimum for $4s(\mathrm{Cu})$ $\zeta = 0.66$ and $1s(\mathrm{H})$ $\zeta = 0.30$. The optimal energy is $E_{\mathrm{Cu}} = -0.3334$.

Taking into account the exact energy of the hydrogen atom, $E_{\mathrm{H}} = -0.5$ a.u., we get

[1]The same internuclear separation was used in our programs for the $f \times g$ approach.

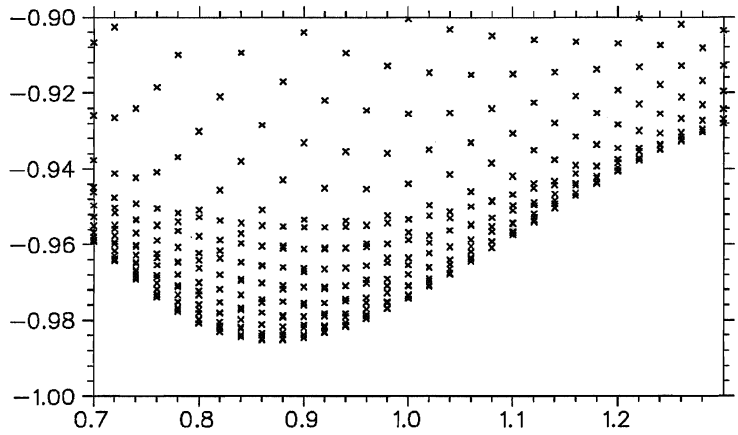

Figure G.1: CuH energy (a.u.) vs. ζ on $4s$ on Cu for a simple SCF model

the dissociation energy $D_e = E_{\mathrm{Cu}} + E_{\mathrm{H}} - E_{\mathrm{CuH}} = 0.1518$ a.u.[2]

This result suggest that even our very simple model of CuH is capable of providing the correct sign for the dissociation energy. However, it is important to realize that this model does not enable us to calculate the internuclear separation, R, and that the observed minimum for the CuH energy is only a local minimum. Both of these limitations are due to the fact that the optimal configuration of a two electron system with two protons is the hydrogen molecule H_2 with the energy -1.175 a.u. and the internuclear separation 1.40 a.u. (see [5]). Therefore, any attempt to find the global minimum of the CuH energy using the presented simple SCF model will lead to mimicking the H_2 molecule.

[2]Experimental value for the dissociation energy is 0.101 a.u. (see table 5.1).

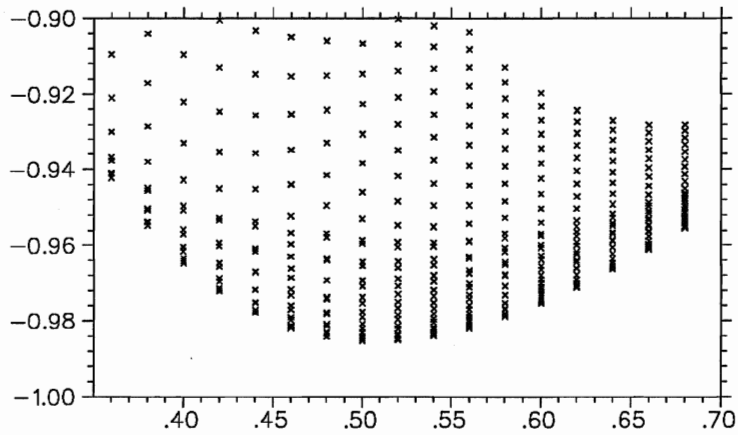

Figure G.2: CuH energy (a.u.) vs. ζ on $1s$ on H for a simple SCF model

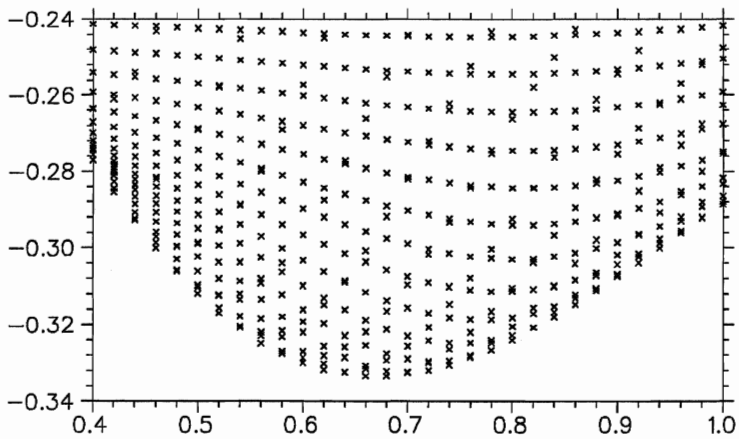

Figure G.3: Cu energy (a.u.) vs. ζ on $4s$ on Cu for a simple SCF model

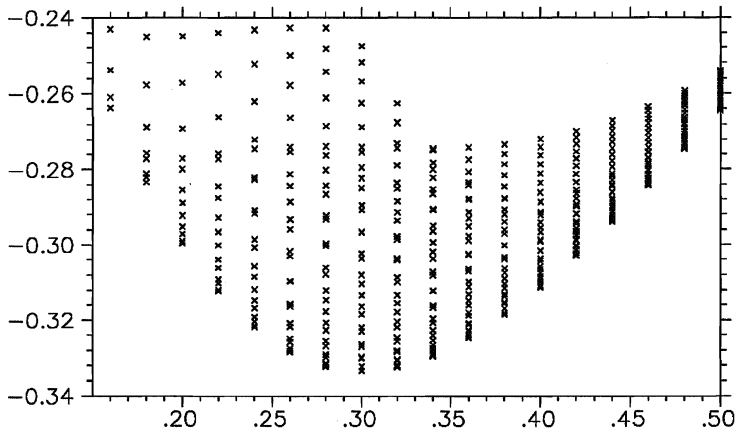

Figure G.4: Cu energy (a.u.) vs. ζ on $1s$ on H for a simple SCF model

Bibliography

[1] Pou-Amérigo, R. et al. *J. Chem. Phys.* **101**, 4893 (1994).

[2] Hammond, B. L., Lester Jr., W. A., and Reynolds, P. J. *Monte Carlo Methods in Ab Initio Quantum Chemistry.* World Scientific Publishing Co. Pte. Ltd., (1994).

[3] Anderson, J. B. *Int. Rev. Phys. Chem.* (1994). (Preprint).

[4] Blinder, S. M. *Am. J. Phys.* **33**, 431 (1965).

[5] Levine, I. N. *Quantum Chemistry.* Prentice-Hall, 4 edition, (1991).

[6] Schaefer, H. F. *Modern Theoretical Chemistry*, volume 3. Plenum, New York, (1977).

[7] Brooks, B. R. and Schaefer, H. F. *J. Chem. Phys.* **70**, 5092 (1979).

[8] Brooks, B. R. et al. *J. Chem. Phys.* **72**, 3837 (1980).

[9] Siegbahn, P. E. M. *J. Chem. Phys.* **72**, 1647 (1980).

[10] Roos, B. O. *Adv. Chem. Phys.* **69**, 399 (1987).

[11] Kraemer, W. P. and Roos, B. O. *Chem. Phys.* **118**, 345 (1987).

[12] Ram, R. S., Bernath, P. F., and Brault, J. W. *J. Mol. Spectrosc.* **113**, 269 (1985).

[13] Armentrout, P. B. and Sunderlin, L. S. *Transition Metal Hydrides.* VCH, New York, (1992).

[14] Pou-Amérigo, R. private consultation.

[15] Salotto, A. W. and Burnelle, L. *J. Chem. Phys.* **52**, 2936 (1970).

[16] Siegbahn, P. E. M. et al. *J. Chem. Phys.* **74**, 2384 (1981).

[17] Roos, B. O. and Taylor, P. R. *Chem. Phys.* **48**, 157 (1980).

[18] Meyer, H. A., editor. *Symposium on Monte Carlo Methods.* Wiley Interscience, New York, (1956).

[19] Metropolis, N. and Ulam, S. M. *J. Am. Stat. Assoc.* **44**, 247 (1949).

[20] Kalos, M. H. *Phys. Rev.* **128**, 1891 (1962).

[21] Kalos, M. H. *Nucl. Phys. A* **126**, 609 (1969).

[22] Kalos, M. H. *J. Comput. Phys.* **1**, 257 (1966).

[23] Kalos, M. H. *Phys. Rev. A* **2**, 250 (1970).

[24] Anderson, J. B. *J. Chem. Phys.* **63**, 1499 (1975).

[25] Anderson, J. B. *J. Chem. Phys.* **65**, 4121 (1976).

[26] Anderson, J. B. *J. Chem. Phys.* **73**, 3897 (1980).

[27] Anderson, J. B. *Int. J. Quant. Chem.* **15**, 109 (1979).

[28] Schulman, L. *Techniques and Applications of Path Integration.* Wiley, New York, (1981).

[29] Chandler, D. *Introduction to Modern Statistical Mechanics.* Oxford, New York, (1987).

[30] Tobochnik, J., Gould, H., and Mulder, K. *Comput. Phys.* **4**, 431 (1990).

[31] Anderson, J. B. and Diedrich, D. L. *J. Chem. Phys.* **100**, 8089 (1994).

[32] Bhattacharya, A. and Anderson, J. B. *J. Chem. Phys.* **100**, 8999 (1994).

[33] Anderson, J. B., Traynor, C. A., and Boghosian, B. M. *J. Chem. Phys.* **99**, 345 (1993).

[34] Anderson, J. B., Traynor, C. A., and Boghosian, B. M. *J. Chem. Phys.* **95**, 7418 (1991).

[35] Vrbik, J. and Rothstein, S. M. *J. Chem. Phys.* **96**, 2071 (1992).

[36] Rothstein, S. M. and Vrbik, J. *J. Comput. Phys.* **63**, 130 (1986).

[37] Hammond, L., Reynolds, P. J., and Lester Jr., W. A. *Phys. Rev. Let.* **61**, 2312 (1988).

[38] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. M., and Teller, E. *J. Chem. Phys.* **21**, 1087 (1953).

[39] Vrbik, J. *J. Phys. A: Math. Gen.* **18**, 1327 (1985).

[40] Vrbik, J. *J. Phys. A: Math. Gen.* **20**, 2693 (1987).

[41] Rothstein, S. M., Vrbik, J., and Patil, N. *J. Comput. Phys.* **8**, 412 (1987).

[42] Umrigar, C. J., Nightingale, M. P., and Runge, K. J. *J. Chem. Phys.* **99**, 2865 (1993).

[43] Wells, B. H. *Methods in Computational Physics*, volume 1. Plenum, New York, (1987).

[44] Hammond, L., Reynolds, P. J., and Lester Jr., W. A. *J. Chem. Phys.* **87**, 1130 (1987).

[45] Ceperley, D. M. *J. Stat. Phys.* **43**, 815 (1986).

[46] Barnett, R. N., Reynolds, P. J., and Lester Jr., W. A. *J. Chem. Phys.* **82**, 2700 (1985).

[47] Reynolds, P. J., Barnett, R. N., Hammond, L., and Lester Jr., W. A. *J. Stat. Phys.* **43**, 1017 (1986).

[48] Mentch, F. and Andreson, J. B. *J. Chem. Phys.* **80**, 2675 (1984).

[49] Garmer, D. R. and Andreson, J. B. *J. Chem. Phys.* **86**, 4025 (1987).

[50] Garmer, D. R. and Andreson, J. B. *J. Chem. Phys.* **86**, 7237 (1987).

[51] Hurley, M. and Christiansen, P. A. *J. Chem. Phys.* **86**, 1069 (1987).

[52] Christiansen, P. A. *J. Chem. Phys.* **88**, 4867 (1988).

[53] Yoshida, T. and Iguchi, K. *J. Chem. Phys.* **88**, 1032 (1988).

[54] Belohorec, P., Rothstein, S. M., and Vrbik, J. *J. Chem. Phys.* **98**, 6401 (1993).

[55] Belohorec, P. Master's thesis, Brock University, (1992).

[56] Bueckert, H., Rothstein, S. M., and Vrbik, J. *Can. J. Chem.* **70**, 366 (1992).

[57] Vrbik, J. private consultation.

[58] Slater, J. C. *Phys. Rev.* **36**, 57 (1930).

[59] Clementi, E. and Roetti, C. *Atomic Data and Nuclear Data Tables* **14**, 177 (1974).

[60] Reynolds, P. J., Ceperley, D. M., Alder, B. J., and Lester Jr., W. A. *J. Chem. Phys.* **77**, 5593 (1982).

[61] Jastrow, R. *Phys. Rev.* **98**, 1479 (1955).

[62] Huber, K. P. and Herzberg, G. *Molecular Spectra and Molecular Structure.* Van Nostrand Reinhold, New York, (1979).

[63] Hobza, P. and Zahradnik, R. *Chem. Rev.* **88**, 871 (1988).

[64] Marian, C. M. *J. Chem. Phys.* **94**, 5574 (1991).

[65] Chong, D. P. et al. *J. Chem. Phys.* **85**, 2850 (1986).

[66] Hirschfelder, J. O. and Nazaroff, G. V. *J. Chem. Phys.* **34**, 1666 (1961).

[67] Fantucci, P., Polezzo, S., Morosi, G., and Valenti, V. *Theoret. Chim. Acta* **67**, 245 (1985).

[68] Jeung, G. H. and Barthelat, J. C. *J. Chem. Phys.* **78**, 2097 (1983).

[69] Stoll, H. et al. *J. Chem. Phys.* **79**, 5532 (1983).

[70] Moskowitz, J. W., Schmidt, K. E., Lee, M. A., and Kalos, M. H. *J. Chem. Phys.* **77**, 349 (1982).

[71] East, A. L. L. Bachelor's thesis, Brock University, (1989).

[72] Ceperley, D. M., Chester, G. V., and Kalos, M. H. *Phys. Rev. B* **16**, 3081 (1977).